Aus dem Institut für Biologie der Humboldt-Universität Berlin

REGULATION OF DNA METHYLATION DURING DEVELOPMENT:

ALTERNATIVE ISOFORMS OF DNA METHYLTRANSFERASE

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

im Fach BIOLOGIE

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I

der Humboldt-Universität zu Berlin

von

Ana Maria Aguirre-Arteta

geb. am 02.01.1968 in Bilbao (Spanien)

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Dr. hc. Hans Meyer

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. Bernhard Ronacher

Gutachter:
1. Prof. Dr. Harald Saumweber
2. PD Dr. Jörn Walter
3. Prof. Dr. Franz Theuring

Tag der mündlichen Prüfung: 28.06.00

REGULATION OF DNA METHYLATION DURING DEVELOPMENT:

ALTERNATIVE ISOFORMS OF DNA METHYLTRANSFERASE

DISSERTATION

To obtain the academic degree

doctor rerum naturalium

(Dr. rer. nat.)

in the field of BIOLOGY

submitted at the

Faculty of Mathematics and Natural Sciences I

Of the Humboldt-University Berlin

von

Ana Maria Aguirre-Arteta

geb. am 02.01.1968 in Bilbao (Spain)

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Dr. hc. Hans Meyer

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. Bernhard Ronacher

Gutachter:
1. Prof. Dr. Harald Saumweber
2. PD Dr. Jörn Walter
3. Prof. Dr. Franz Theuring

Tag der mündlichen Prüfung: 28.06.00

THIS WORK IS DEDICATED WITH LOVE TO MY PARENTS, MY BROTHER AND SPECIALLY TO MY SISTER AMAYA

DIE WISSENSCHAFT FAENGT EIGENTLICH ERST DA AN INTERESSANT ZU WERDEN, WO SIE AUFHOERT

JUSTUS VON LIEBIG (1803-1873)

APPENDIX

A

A:	adenine
Ab:	antibody
A_{260}:	absorbance at 260 nm
A_{280}:	absorbance at 280 nm
Amp:	ampicillin
AMV:	avian myeloblastosis virus
AP:	alkaline phosphatase
APS:	ammonium persulphate
Arg:	arginine
Asp:	aspartate
ATP:	adenosine triphosphate

B

B:	BamHI
bp:	base pair
BSA:	bovine serum albumin

C

C:	cytosine
cDNA:	complementary DNA
COS-7 cells:	african green monkey kidney fibroblast-like cells transformed with SV40 T antigen
CpG:	cytosine-guanine doublet
cpm:	counts per minute
CTP:	cytosine triphosphate
Cys:	cysteine

D

D:	diplotene
Da:	dalton
dATP:	deoxyadenosine triphosphate
dCTP:	deoxycytidine triphosphate
ddNTP:	2', 3'-dideoxynucleotide-5'-triphosphate
dGTP:	deoxyguanosine triphosphate
DMPC:	dimethylpyrocarbonate
DMSO:	dimethyl sulfoxide

dMTase:	DNA demethylase
DNA:	deoxyribonucleic acid
DNA MTase:	DNA methyltransferase
DNase:	deoxyribonuclease
dNTP:	2'-deoxynucleotide-5'- triphosphate
dpc:	day post coitum
DTT:	dithiothreitol
dTTP:	deoxythymidine triphosphate
	E
E:	EcoRI
EDTA:	ethylenediaminetetraacetic acid
ES cells:	embryonic stem cells
	F
F:	forward
FCS:	fetal calf serum
FITC:	fluorescein isothiocyanate
	G
G:	guanine
GAPDH:	glyceraldehyde phosphate dehydrogenase
Gln:	glutamine
Glu:	glutamate
Gly:	glycine
g:	gram
GTP:	guanosine triphosphate
	H
h:	hour
HDAC:	histone deacetylase
HhaI:	*Haemophilus haemolyticus*
Hoechst 33258:	2' (4-hydroxyphenyl)-5-(4-methyl-1-piperazinyl)-2,5' bi-1H-benzimidazole
HRP:	horseradishperoxidase
	I
IgG:	immunoglobulin G
In:	intermediate spermatogonia

IPTG:	isopropylthio-β-D-galactoside
	K
kb:	kilobase
kDa:	kilo Dalton
	L
L:	leptotene
LB:	Luria Bertani
Lys:	lysine
	M
M:	molar (mol/l)
mA:	milliampere
MB:	myoblast
5 mC:	C5-methylcytosine
MDG1:	microvascular endothelial differentiation gene 1
MeCP:	methyl-CpG-binding protein
mg:	milligram
min:	minute
ml:	milliliter
mmol:	millimol
MoMuLV:	moloney murine leukemia virus
MOPS:	3-(N-morpholino) propanesulfonic acid
mRNA:	messenger RNA
MT:	myotube
MTase:	methyltransferase
	N
NFκB:	NF-kappa B
ng:	nanogram
NLS:	nuclear localization signal
nm:	nanometer
	O
OD:	optical density
ORF:	open reading frame
	P
P:	pachytene

PAGE:	polyacrylamide electrophoresis
PBHD:	polybromo-1 protein homologous domain
PBS:	phosphate buffer saline
PCNA:	proliferating cell nuclear antigen
PCR:	polymerase chain reaction
PGC:	primordial germ cell
pH:	negative log of hydrogen ion concentration
Phe:	phenylalanine
PI:	preleptotene
pRb:	retinoblastoma protein
Pro:	proline
PVDF:	polyvinylidenfluorid
pVHL:	von Hippel-Lindau tumor suppressor protein
Pwo:	Pyrococcus woesei

R

R:	reverse
RACE:	rapid amplification of cDNA ends
RFTS:	replication foci targeting sequence
RNA:	ribonucleic acid
RNase:	ribonuclease
rRNA:	ribosomal RNA
RT:	reverse transcription

S

S:	SalI
SAH/AdoHCy:	S-adenosyl-L-homocysteine
SAM/AdoMet:	S-adenosyl-L-methionine
SDS:	sodium dodecyl sulfate
Ser:	serine
SSC:	standard saline citrate
SV40:	simian virus 40

T

T:	thymine
TAE:	tris acetate EDTA
Taq:	*Thermus aquaticus*

TBS:	tris buffered saline
t-exon:	testis specific exon
TE:	tris EDTA
TEMED:	N, N, N', N'-tetramethyl-ethylendiamine
Tf:	transcription factor
Tm:	melting temperature
TRD:	target recognition domain
Tris:	tris (hydroxymethyl)-aminomethane
tRNA:	transfer RNA
TTP:	thymidine triphosphate
Tyr:	tyrosine
	U
U:	uridine
UTP:	uridine triphosphate
UV:	ultraviolet light
	V
V:	volt
	X
X:	Xbal
X-gal:	5-bromo-4-chloro-3-indolyl- -D-galactopyranoside
	Z
Z:	zygotene
Zn:	Zinc
	µ
µCi:	microCurie
µg:	microgram
µl:	microliter

DNA methyltransferases (DNA MTases) are enzymes responsible for DNA methylation (transfer of methyl groups to a base in the DNA) and are vital for the development of mammals. Several MTases have been identified in eukaryotes but the most abundant is Dnmt1. Furthermore, many pathological conditions are often attributed to an altered availability or function of this enzyme, however the understanding of the regulation of Dnmt1 and the concomitant relationship to diseases is far from being complete. In mammals the methylation of DNA correlates with gene activity, and methylation patterns change dramatically during early development when the genome of the mammalian embryo undergoes consecutive waves of demethylation (loss of methylation) and *de novo* methylation (methylation of DNA sites that have not been previously methylated). The hypothesis of this study was that alternative Dnmt1 isoforms are expressed at specific developmental stages and thus contribute to changes in the DNA methylation pattern. To study this regulation the structure of the Dnmt1 gene was determined. In this work, the tissue distribution and abundance of Dnmt1 mRNA was analyzed by Northern blot and a new, longer transcript was identified that is present in testis and skeletal muscle tissue. The novel isoform was cloned by a combination of RT-PCR and RACE techniques and found to be identical in both tissues. This new isoform differs from the ubiquitous cDNA in the 5' end, utilizing a new transcriptional start site and an 800 bp long alternative first exon. The cellular localization of this new transcript was determined by in situ hybridization and found to be present in the more specialized haploid spermatogenic cells, spermatids and at lower level in skeletal muscle. During muscle differentiation, the ubiquitous isoform is downregulated while the alternative isoform is upregulated. Although this mRNA codes for several short upstream ORFs which could prevent translation of the Dnmt1-specific ORF, it was found by immunofluorescence and Western blot analyses that this transcript can be translated *in vivo* producing a shorter Dnmt1 protein. The results shown here indicate that alternative Dnmt1 isoforms are expressed *in vivo* and might play an active role in the regulation of DNA methylation.

Die DNA Methyltransferasen sind verantwortlich für die spezifische Methylierung von DNA-Basen. Mehrere DNA Methyltransferasen sind bekannt, wobei die Dnmt1 das hauptsächlich vorkommende Enzym ist. Bei Säugetieren korreliert die DNA-Methylierung mit der Genaktivität und ist essentiell für die Embryonalentwicklung. Eine beeinträchtigte Funktion oder Verfügbarkeit des Enzyms kann zu pathologis
führen. Die Regulation der Dnmt1 und die damit verbundene Bedeutung bei der Entstehung von Krankheiten ist bisher nur unvollständig untersucht. In der Frühphase der Embryonalentwicklung von Säugetieren ändert sich das Methylierungsmuster des Genoms dramatisch. In zeitlich aufeinander folgenden Phasen wird die DNA demethyliert (Verlust der Methylgruppen) und neu methyliert (De-Novo Methylierung). Die Hypothese dieser Arbeit ist, dass verschiedene Isoformen der Dnmt1 in spezifischen Entwicklungsstadien exprimiert werden und zu Veränderungen des Methylierungsmusters der DNA beitragen. Um diese Regulation zu untersuchen, wurde die Struktur der Maus Dnmt1-Gens bestimmt. Außerdem wurde in verschiedenen Gewebetypen die Transkriptionsgröße und die Transkriptionsintensität der mRNA mit Hilfe von Northern-Blots quantifiziert. Mit diesen Experimenten konnte im Hoden- und Skelettmuskelgewebe ein längeres Dnmt1 Transkript als in anderen Geweben identifiziert werden. Dieses neue Dnmt1-Transkript wurde mit Hilfe von RT-PCR und RACE-Techniken kloniert und ist in beiden Geweben identisch. Es unterscheidet sich auf DNA-Ebene in der Sequenz des 5'-Endes von der bisher bekannten Form der Dnmt1 und besitzt einen anderen Startpunkt für die Transkription. Darüber hinaus besitzt das neue Dnmt1-Transkript ein 800 Basenpaar großes erstes Exon, welches sich von dem des bekannten Dnmt1-Transkripts unterscheidet. Die spezifische zelluläre Lokalisation des neuen Transkripts wurde mit Hilfe der In-Situ-Hybridisierung analysiert. Mit dieser Technik wurde das alternative Transkript in stärker spezialisierten, haploiden spermatogenen Zellen (Spermatiden) und zu einem geringen Maß im Skelettmuskel nachgewiesen. Während der Differenzierung von Muskelzellen wurde eine verminderte Expression des bereits bekannten mRNA-Transkripts und eine verstärkte Expression des neu identifizierten mRNA-Transkripts festgestellt. Obwohl die mRNA der alternativen Isoform verschiedene, kurze offene Leserahmen enthält, welche die Translation eines spezifischen Dnmt1 Proteins verhindern könnten, wurde durch Immunofluoreszenz- und Western-Blot Analysen ein Translationsprodukt nachgewiesen. Nach den hier aufgezeigten Ergebnissen werden alternative Dnmt1 Isoformen *in vivo* exprimiert, welche eine aktive Rolle bei der Regulation der DNA-Methylierung spielen könnten.

1.1 DNA METHYLATION

The DNA of most organisms has in addition to the four major bases; A, T, C and G, one or more minor bases. These minor bases are created by a modification of the DNA by post-replicational methylation (Riggs, 1990).

1.1.1 TYPES OF DNA METHYLATION

Two major classes of DNA modifications exist in nature. One class methylates a ring carbon in position 5 to form C5-methylcytosine (5mC). The second class methylates exocyclic nitrogens and forms either N4-methylcytosine or N6-methyladenine.

In prokaryotes, the major role of DNA methylation is to protect host DNA against degradation by the cognate restriction enzymes. However, *Escherichia coli* also has two methylase systems that are not associated with restriction-modification systems. The *dam* gene encodes a methylase that forms 6mA and the *dcm* gene encodes a methylase that forms 5mC. Methylation has been shown not to be essential in *Escherichia coli* (Russell and Hirata, 1989), but provides useful functions. The classic function is the DNA modification in a restriction-modification system. Foreign DNA from a species or strain with a different or no methylation pattern is digested by the restriction enzyme of the host restriction-modification system. Some strains of *Escherichia coli* (e.g. K-12) are able to restrict methylated DNA protecting it against infections by phages which escape the conventional host restriction by methylating their DNA while they replicate in the host strains. Therefore, DNA methylation, along with restriction enzymes, function to reduce the efficiency of gene transfer between unrelated species. The dam methyltransferase (methylase or MTase) is involved in a wide variety of cellular functions such as postreplicative mismatch repair, control of *Escherichia coli* chromosome replication and chromosome segregation, control of plasmid replication, transposition, gene expression and control of the initiation of phage P1 DNA packaging (Marinus, 1984).

In eukaryotes there is no evidence for a restriction enzyme function. All potential sites are methylated in bacteria but many potentially methylatable sites are unmethylated in eukaryotes. Unmethylated sites would be sensitive to restriction if restriction modification systems similar to those in bacteria were in function. Therefore, it is likely that methylation serves other functions in eukaryotes. In these organisms, DNA methylation has been implicated in the control of several cellular processes, mainly in gene regulation and thus in processes including differentiation and development. C5 cytosine MTases can be found in both prokaryotes and eukaryotes, whereas the exocyclic MTases have been mainly isolated from prokaryotes. In general, the 5mC modification is nearly universal among eukaryotes with genomes greater than 10^8 bp (mammals and plants), but rare among eukaryotes with smaller genomes (yeast, flies and nematodes) (Bestor, 1990; Bird, 1992). Furthermore, two C5 cytosine MTases (masc1 and masc2) with similarities to the mammalian MTases have been cloned from the filamentous fungus *Ascobolus immersus* (Chernov et al., 1997; Malagnac et al., 1997). DNA methylation in *Ascobolus* is essential for the completion of the developmental processes underlying sexual reproduction. To date, 5mC has not been detected in the DNA of *Schizosaccharomyces pombe* (Antequera et al., 1984), *Saccharomyces cerevisiae* (Proffitt et al., 1984), *Caenorhabditis elegans* (Simpson et al., 1986) and *Drosophila melanogaster* (Urieli-Shoval et al., 1982). Recently a MTase related *Drosophila* protein has been cloned, DmMT2, (Hung et al., 1999) which is similar to that of vertebrate Dnmt2 enzymes. This work focuses on the regulation of DNA methylation by the C5 cytosine MTases in mammals.

1.1.2 DNA METHYLATION REACTION

The enzyme responsible for the methylation process is DNA MTase, which transfers a methyl group from S-adenosyl-L-methionine (SAM or AdoMet) to the 5-carbon in the pyrimidine ring of cytosine (Fig.1.1) (Wu and Santi, 1987). This process involves the formation of a transient covalent complex between the protein and the pyrimidine being modified (see Fig. 1.1). A cysteine thiol of the enzyme attacks the 6-carbon of cytosine and forms a covalent DNA-protein intermediate. The addition of the cysteine thiol activates the 5-carbon allowing transfer of the methyl group from SAM and release of S-adenosyl-L-homocysteine (SAH or AdoHcy).

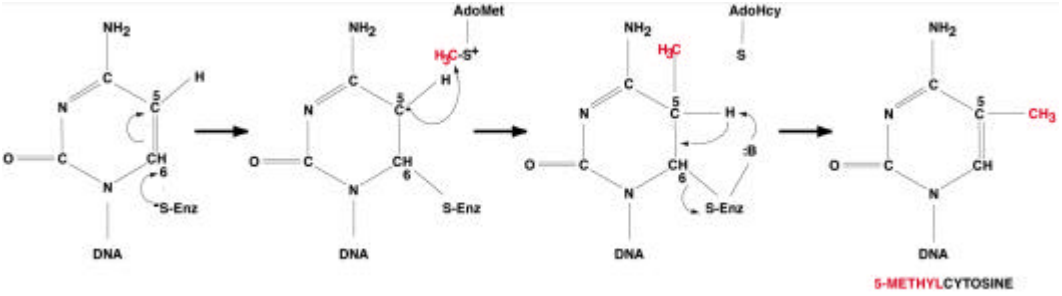


Fig.1.1. Reaction pathway for C5-cytosine methyltransferase based on the mechanism proposed by Wu and Santi (Wu and Santi, 1987) ***for thymidylate synthase and tRNA-(uracil-5)methyltransferase***. *The attack on carbons 5 and 6 are shown in the figure. The methyl group is transferred to carbon-5 of cytosine, and the hydrogen located on carbon-5 is released as H^+. During this reaction SAM is coverted to SAH. A base (B:) that abstracts a proton from carbon-5 is needed for the elimination step* (Gerlt and Gassman, 1993).

The fact that DNA contains a set of modifications that is not encoded in the genetic sequence but is related to changes in gene activity, classifies DNA methylation as an epigenetic process. The recognition sequence for the mammalian DNA MTase has also been identified and is highly specific with almost all cytosine methylations occurring on 5'-C-p-G-3' (CpG) (Gruenbaum et al., 1981) and over half of CpG dinucleotides being methylated in the mammalian genome (Bestor et al., 1984). It has also been found that cytosine methylation can occur at CpXpG trinucleotides in higher plants (Gruenbaum et al., 1982) and also in mammals but at lower frequency (Clark et al., 1995).

1.1.3 STRUCTURE OF A C5-CYTOSINE METHYLTRANSFERASE

Several aspects of the catalytic mechanism of DNA methylation have been clarified mainly for prokaryotic MTases (Wu and Santi, 1987). MTase HhaI is a C5 cytosine MTase originally isolated from the bacterium *Haemophilus haemolyticus* (Roberts et al., 1976). The gene for this MTase has been cloned and sequenced and the overexpression of the protein in *Escherichia coli* (Caserta et al., 1987; Klimasauskas et al., 1990) followed by large-scale purification has allowed its crystallization (Kumar et al., 1992). The enzyme is folded into two domains (Fig.1.2): a larger catalytic domain and a smaller DNA recognition domain which is involved in flipping the target cytosine out and

positioning it into the catalytic pocket of the enzyme (Cheng, 1995) (Fig.1.2).

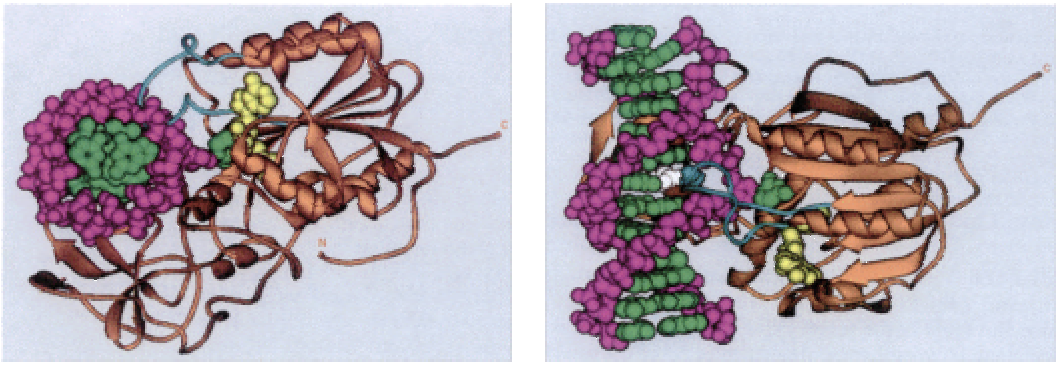


Fig.1.2. Graphic representation of the complex of the MTase HhaI covalently bound to a 13-mer DNA duplex containing its recognition sequence. The end product of the reaction, SAH, is also present (yellow). *The protein is shown in brown, the DNA backbone in magenta, the DNA bases in green, and the active-site loop and the recognition loops are represented in white (right panel). The left panel is a view looking down the DNA helix axis, in which the DNA can be seen to lie between the large domain (on the right) and the small domain (on the left). On the right panel a side view from the minor groove is shown. (Fig. from (Cheng, 1995)).*

1.1.4 DNA METHYLATION AND GENE EXPRESSION

The methylation of DNA helps to regulate gene expression and mistakes in this regulation can have drastic consequences leading to birth defects and cancer. Three possible ways by which DNA methylation can affect gene expression are:

i) **Direct effect:** Methylated CpG residues can directly interfere with the binding of specific transcription factors to DNA. Several transcription factors (AP-2, c-Myc/Myn, E2F and NFκB) bind to DNA sequences that include CpG sites and have been shown to be sensitive to methylation at these sites (TF1 in Fig. 1.3). Other transcription factors, however, are not sensitive to methylation (Sp1, CTF and YY1) (Tate and Bird, 1993) (TF2 in Fig. 1.3).

ii) **Direct binding of specific factors:** The direct binding of specific factors to methylated DNA mediates repression (Fig. 1.3). Two such factors, MeCP1 (complex of several subunits) and MeCP2, have been identified and shown to bind to methylated CpG in any sequence context and thereby preventing transcription factor binding. MeCP1 binds to DNA containing multiple symmetrically methylated CpG sites (Meehan et al., 1989). MeCP2 is more abundant than MeCP1 and is able to bind to DNA that is asymmetrically methylated, with just a single methyl-CpG. In addition, the distribution of MeCP2 on the chromosome parallels that of methyl-CpG (Lewis et al., 1992). Experiments on mice with a disrupted MeCP2 gene have shown that MeCP2, like DNA MTase is dispensable in embryonic stem cells but essential for embryonic development (Tate et al., 1996).

iii) **Chromatin structure alteration:** It has been shown that methylated DNA affects the positioning of nucleosomes and influences the sensitivity of DNA to DNAase I (Keshet et al., 1986). Further experiments using microinjection of methylated and non-methylated templates into nuclei have shown that methylation inhibits expression only after chromatin is assembled (Buschhausen et al., 1987). These results support the view that methylation induces a change in conformation of chromatin to an inactive state (Kass et al., 1997). However, only recently it has been shown how cytosine methylation affect the structure of chromatin (Fig. 1.3) by its interaction with MeCP2 (Jones et al., 1998; Nan et al., 1998). MeCP2 binds to chromosomes in a methylation dependent manner and mediates transcriptional repression. MeCP2 contains a transcriptional repression domain which associates with a transcriptional repression complex containing mSin3a and histone deacetylase (Jones et al., 1998; Nan et al., 1998). Two possible explanations of how this interaction works have been proposed. In the first one deacetylation of lysine amino groups might allow interactions between the amino terminal histone tails and the DNA phosphate backbone. The second possibility is that deacetylation might favour interactions between adjacent nucleosomes (Fig.1.4).

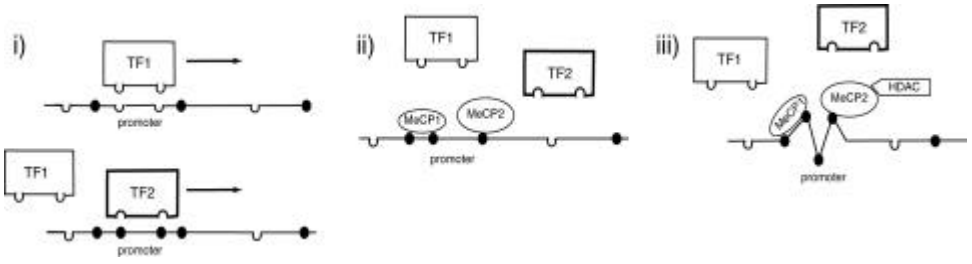
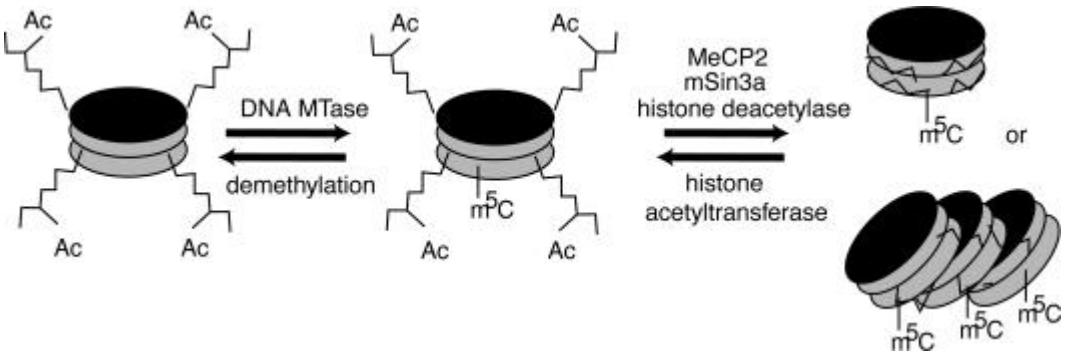


Fig.1.3. A model to explain the effects of DNA methylation on transcription*. The relevant parameters are proximity and methylation of CpG sites. The MeCP molecules are shown as ovals, the CpG sites as filled circles (methylated) or half circles (non-methylated) circles, the arrows represent transcription from the promoter, TF1 and 2 represent transcription factors 1 and 2, sensitive or insensitive to methylation, respectively and HDAC represents the histone deacetylase. Active transcription is indicated with an arrow.*

Fig.1.4. Methylation and histone deacetylation*. MeCP2, a protein that binds to methylated DNA, exists in a complex with histone deacetylase and mSin3a. Nucleosomes (as several connecting disks) with their histone tails (as thin lines coming out of the nucleosomes) are shown. The two mechanisms in which histone deacetylation changes the structure of the nucleosomes are also depicted. The deacetylation by the MeCP2 œomplex (MeCP2, mSin3a and histone deacetylase) of the lysine amino groups on the histone tails might allow the interaction of the histone tails and the DNA backbone or might lead to compaction of the chromatin by allowing the interactions between the nucleosomes (Figure from (Bestor, 1998)).*



1.2 DNA METHYLATION IN NORMAL DEVELOPMENT AND DISEASE

1.2.1 DNA METHYLATION AND DEVELOPMENT

A distinct change in the degree of methylation of the genome has been observed during gametogenesis and early mammalian embryogenesis. While many genes are highly methylated in sperm DNA (Groudine and Conkin, 1985; Razin et al., 1984), genes are often less methylated in oocytes (Monk et al., 1987; Sanford et al., 1987) (Fig. 1.5). During development, the DNA of the extraembryonal membranes (yolk sac and placenta) becomes dramatically demethylated, while the DNA of fetal tissues is subjected to a *de novo* methylation process after implantation (Monk et al., 1987; Razin et al., 1984; Sanford et al., 1987).

i) **DNA methylation in the early embryo.** Dynamic changes in the methylation pattern of specific genes in the early embryo were observed. CpG islands are regions (500-2000 bases pairs) of unmethylated DNA with a high frequency of CpG dinucleotides (Bird et al., 1985) and have been found to be associated with 5' ends of housekeeping genes and of some tissue specific genes (Bird, 1986). The CpG islands in housekeeping genes are unmethylated throughout development. However, the CpG islands in tissue-specific genes are heavily methylated in sperm and less methylated in the mature oocytes. Between the 4 to 16 cell stages this methylation is erased. The unmethylated state of these genes persists throughout the entire preimplantation period (Howlett and Reik, 1991; Kafri et al., 1992; Sanford et al., 1987).

ii) **Methylation during gametogenesis.** Primordial germ cells (PGCs) emerge from the generally unmethylated epiblast before tissue specific de novo methylation takes place in the epiblast lineage, which gives rise to the embryo proper (somatic lineage). The trophoblast derived lineage (placenta and yolk sac) is substantially demethylated. The PGCs then migrate and colonize the developing genital ridges by 10-11 dpc (Gomperts et al., 1994). At 12.5 dpc before sexual differentiation of the primordial germ cells begins, tissue specific and imprinted genes are unmethylated (Chaillet et al., 1991; Kafri et al., 1992). This hypomethylated state prevails until sexual differentiation of the PGCs takes place

and *de novo* methylation occurs in the cells populating the gonads of both sexes, forming a bimodal pattern of methylation where CpG islands remain unmethylated and non-island sequences are methylated (Kafri et al., 1992).

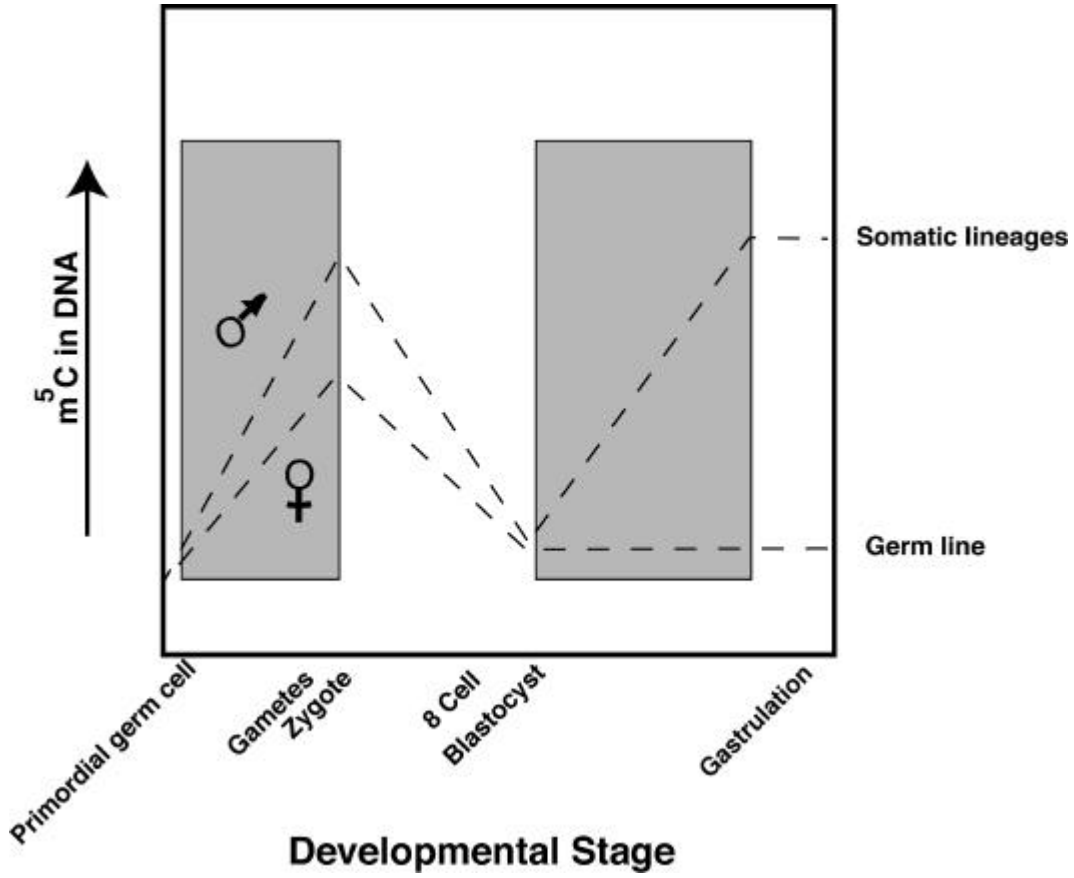


Fig.1.5. Methylation during gametogenesis and early development in mammals. *The overall methylation increases during gametogenesis being higher in the male gametes. During cleavage, the genome undergoes global demethylation and after the blastocyst stage the overall methylation level remains low in the germ line and increases in the somatic lineage. There are therefore two waves of de novo methylation (grey boxes), one is during gametogenesis and the other is after implantation when tissue specific patterns are being established (Fig. modified from T. H. Bestor).*

This work is concerned with the regulation of DNA methylation during gametogenesis and myogenesis and focusses on the identification and characterization of alternative *Dnmt1* isoforms.

1.2.2 DNA METHYLATION AND DISEASE

Global changes in methylation levels as well as modification of methylation patterns of individual genes has been observed in a variety of tumors (Cardoso and Leonhardt, 1999). However, the role of methylation in carcinogenesis is still not clear since both DNA

hypomethylation and hypermethylation have been associated with cancer (Baylin et al., 1991; Jones, 1986).

i) **Epigenetic changes induced by the (cytosine-5)-DNA MTase:** The mRNA level and enzyme activity of MTase are higher in many tumor cells than in normal cells (Kautiainen and Jones, 1986). However, the generally higher methyltransferase activity does not lead to an overall increase but rather to a decrease of total genomic 5mC content (Feinberg et al., 1988; Gama-Sosa et al., 1983). When the methylation status of various genes is measured in tumor cells, both hypermethylation as well as hypomethylation become apparent (Counts and Goodman, 1995). It appears that the pattern of methylation of specific genes is altered in tumor cells and normally unmethylated CpG islands may become hypermethylated (Baylin et al., 1991).

Hypomethylation. Reduced levels of global DNA methylation have been found in a variety of malignancies. Growth inducing genes such as oncogenes can become overexpressed as a result of hypomethylation. In particular, demethylation and overexpression of the the c-fos, the c-myc and the c-H-ras proto-oncogenes are known to be involved in hepatocarcinogenesis induced by conditions such as methyl donor starvation (Rao et al., 1989; Wainfan and Poirier, 1992).

Hypermethylation. There have been reports of regional increases in DNA methylation levels and regional hotspots for hypermethylation on chromosomes 3p, 11p and 17 p in a variety of human tumors have been shown (de Bustros et al., 1988; Makos et al., 1993). Genes involved in growth arrest and terminal differentiation such as the *MyoD1* gene (Rideout et al., 1994) and tumor suppressor genes such as pRb (Greger et al., 1989; Sakai et al., 1991), bcr-abl (Zion et al., 1994), pVHL (Herman et al., 1994), pWT (Steenman et al., 1994), estrogen receptor (Lapidus et al., 1996), $p57^{KIP2}$ (Hatada et al., 1996), MDG1 (Huynh et al., 1996) and p16 (Little and Wainwright, 1995) are often found inactivated by either mutation or hypermethylation in tumor cells. Hypermethylation of tumor suppressor genes may lead to gene inactivation and result in a selective growth advantage of affected cells.

ii) **Genetic changes at the target site of (cytosine-5) DNA MTase:** Cytosine methylation is also responsible for the induction of a high percentage of disease-causing point mutations in tumor suppressor genes in somatic and germline cells. Mutations occur by spontaneous deamination of 5mC (Coulondre et al., 1978). Both cytosine (C) and 5mC deaminate in single- and double-stranded DNA to form uracil and thymine, respectively (Lindahl, 1993). However, it is more difficult for the cell to correct the resulting T:G mismatch since thymine, unlike uracil, is a normally occurring DNA base. In aqueous solution at 37^{o}C, a 4- to 9- fold higher deamination frequency of 5-mC relative to C in single stranded DNA and in double stranded DNA a ~ 2-fold higher deamination frequency was measured for 5mC when compared to C (Shen et al., 1994). A sophisticated DNA repair apparatus has evolved in mammals which catalyzes the selective repair of the T-G mismatches (resulting from 5mC deamination) back to C-G pairs (Wiebauer and Jiricny, 1989; Wiebauer and Jiricny, 1990). The remaining unrepaired C to T transitions cause many gene mutations in humans leading to hereditary diseases and cancer (Green et al., 1990; Rideout et al., 1990).

Vertebrates and other organisms with DNA methylation show a depletion in the frequency of occurrence of the CpG dinucleotide, which is the predominant site at which 5mC is found. This depletion is thought to be due to increased mutagenesis as described above.

The target cytosines of MTase in prokaryotes are mutated to thymine with an increased frequency when compared with non target cytosines. In eukaryotes, the enhanced mutability of 5mC contributes significantly to carcinogenesis and inherited disease (Jones et al., 1992; Laird and Jaenisch, 1994). In humans, several genes such as p53 (Greenblatt et al., 1994; Tornaletti and Pfeifer, 1995) and p16 (Pollock et al., 1996) are frequently mutated at CpGs in tumor cells and 30% of all inherited mutations are believed to occur by mutations at CpG sites in the germ line (Cooper and Youssoufian, 1988; Jones et al., 1992; Laird and Jaenisch, 1994).

iii) Metabolism of the cofactor SAM and its involvement in carcinogenesis: Several studies indicate that aberration of the metabolism of the cofactor SAM may also play an important role in carcinogenesis (Chiang et al., 1996). Aberrations of the SAM metabolism may not only affect DNA methylation but also other methyl-transfer reactions (e.g. synthesis of spermine and spermidine). Insufficient supply of methionine, folate and choline leads to hypomethylation of liver DNA, increased expression of the *c-H-ras, c-jun and c-myc* genes and to the generation of liver tumors in rats (Simile et al., 1994; Wainfan and Poirier, 1992). In humans, methyldonor deficiency is correlated with an increased risk for liver and colon tumors (Giovannucci et al., 1993).

1.2.3 DNA METHYLATION AND IMPRINTING

Imprinted genes are those genes whose expression is determined by their parental origin, and it has been reported that cancer cells sometimes show loss of imprinting (Feinberg, 1993) which might be explained by changes in DNA methylation that are known to accompany tumorigenesis. Other diseases that have been reported to be associated with imprinting include Prader-Willi (Nicholls et al., 1998), Angelman (Nicholls et al., 1998) and Rett (Willard and Hendrich, 1999) syndromes.

Furthermore, X-inactivation in female mammals is an event associated with a methylation change in CpG islands. The CpG islands within the promoter regions of the X-linked housekeeping genes become highly methylated at or soon after gene inactivation. The methylation of CpG islands on inactive X-chromosome has important medical implications with regard to the fragile X-chromosome, and it has been found that a CpG island near to the breakage site is methylated in most of the affected people (Bell et al., 1991). In contrast to X-linked genes, the CpG islands associated with autosomal genes remain unmethylated, with some exceptions. These exceptions include Alu elements (Kochanek et al., 1993) and L1 retrotransposons (Crowther et al., 1991) which are frequently methylated in the human genome. Although methylation is inherently mutagenic and has led to genomic alterations, it is tolerated possibly because of its contribution to the regulation of gene expression and its necessity during embryonic development.

1.3 REGULATION OF DNA METHYLATION

DNA methylation patterns are known to be maintained over multiple cell generations, however, are not the same in different cell types. The particular patterns are established during early development and can change during development and disease. A model to explain the maintenance and change of cell specific patterns of methylation has been presented (Holliday and Pugh, 1975; Razin and Riggs, 1980; Riggs, 1975). A maintenance MTase was postulated that has a strong preference for hemimethylated sites and thus maintains a given methylation pattern after DNA replication. In addition, multiple *de novo* MTases were postulated, that are expressed and active at specific stages during development.

i) **Maintenance of methylation.** Replication results in the generation of a hemimethylated double-stranded DNA, composed of a methylated parental strand and unmethylated nascent strand. Only this hemimethylated DNA serves as a substrate for the maintenance MTase (Fig. 1.6). This model proposes that DNA methylation patterns are restored after replication because DNA MTase is very efficient in methylating hemimethylated CpG sites (maintenance methylation) but relatively inefficient in methylating nonmethylated substrates.

ii) ***De novo* methylation.** DNA MTases can also methylate certain unmethylated CpG sites that are not in a hemimethylated configuration, a process referred to as *de novo* methylation (Fig.1.6). *De novo* methylation occurs extensively during gametogenesis and after implantation (Fig1.5) (Monk et al., 1987). Several studies have shown *de novo* methylated cytosines in genomic regions containing preexisting methylated cytosines (methylation spreading), such as it occurs at sites where viral DNA integrates in the genome (Doerfler et al., 1990; Riggs, 1990; Toth et al., 1990; Turker, 1990). Since cytosine methylation can directly and indirectly affect the DNA binding of certain transcriptional regulatory factors (Fig. 1.3), the introduction of additional methylated cytosines within gene regulatory sequences may influence gene expression (Doerfler et al., 1990). This spreading of cytosine methylation in gene regulatory sequences has been implicated in the gene silencing characteristic of fragile X syndrome (Oberle et al., 1991; Pieretti et al., 1991), cellular senescence (Tollefsbol and Andrews, 1993), and X-chromosome inactivation (Pfeifer et al., 1990).

iii) **Demethylation.** Demethylation may occur throughout development but is more pronounced during preimplantation development. Demethylation processes have also been shown during gene activation in resting cells (Jost and Jost, 1994). Two possible pathways have been described for the loss of DNA methylation:

Passive demethylation: This occurs by a failure to methylate newly replicated DNA (Razin and Riggs, 1980) (Fig. 1.6). This type of `demodification´ takes place after inactivation of cellular MTases with 5-azacytidine (Jones, 1984). Furthermore, it has been shown that a correlation exists between the overall decrease in the genomic methylation level and the active retention of Dnmt1 in the cytoplasm from the oocyte to the blastocyst stage (Cardoso and Leonhardt, 1999; Carlson et al., 1992).

Active demethylation: It has been shown that methyl groups are removed when DNA is not even undergoing replication, such as that seen on inducing the vitellogenin gene in the chick liver (Wilks et al., 1984) or the globin gene in stimulated erythroleukemia cells (Razin et al., 1986). Active demethylation has been described in transiently transfected myoblasts (Paroush et al., 1990), in postmeiotic spermatocytes (Trasler et al., 1990) and in preimplantation mouse embryos (Kafri et al., 1993). This active process of demethylation could involve 5mC DNA glycosylase activity which has been detected in HeLa nuclear extracts (Vairapandi and Duker, 1993). Other studies have also shown that the demethylation process is at least partly mediated by RNA molecules and that demethylation takes place through the removal of DNA nucleotides and their conversion to an RNase-sensitive form (Weiss et al., 1996). Recently, a human cDNA that encodes a DNA demethylase (dMTase) activity that can catalyze the removal of a methyl residue from 5mC and its release as methanol has been cloned (Bhattacharya et al., 1999).

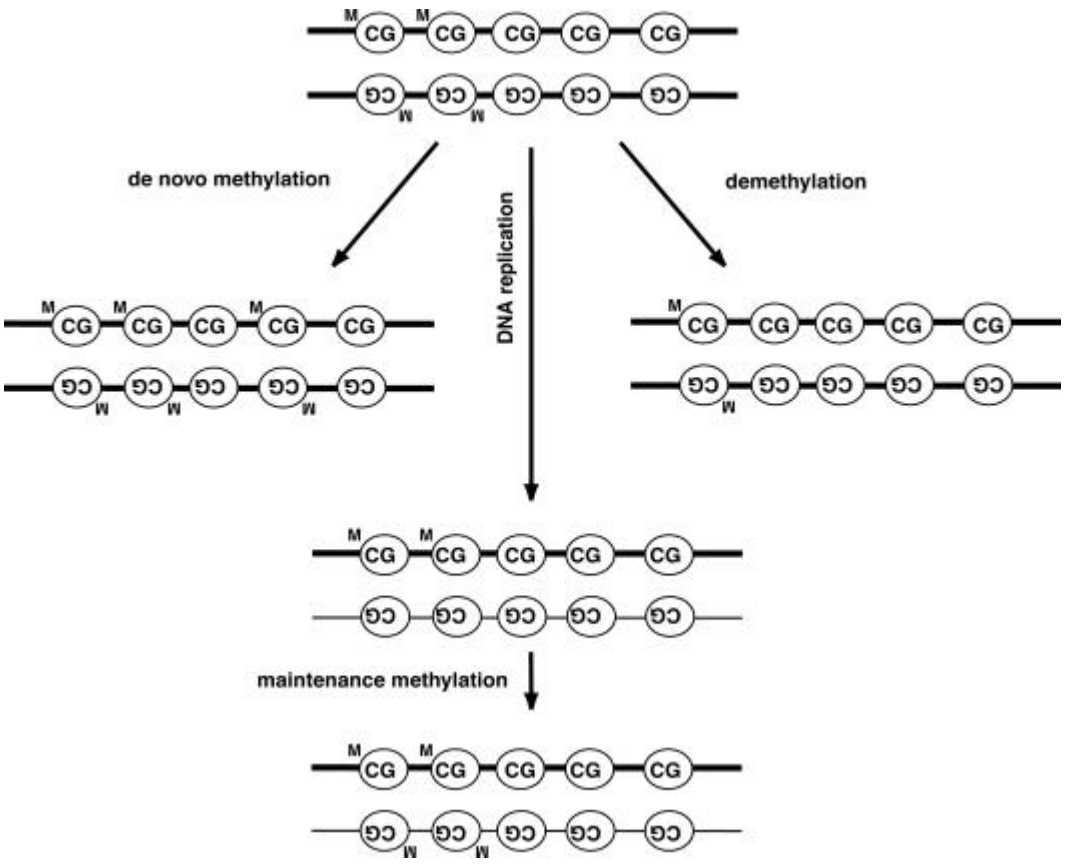


Fig.1.6. Establishment and maintenance of DNA methylation patterns. *The establishment of DNA methylation patterns involves a number of processes. Left, de novo methylation: when a potential methylatable site (indicated as a circle) that is not methylated on both strands of the DNA undergoes methylation (indicated as M outside the circles). Center, maintenance methylation: the DNA molecule is composed of a parental methylated strand (thick line) and a nascent unmethylated strand (thin line). Maintenance DNA MTases add methyl groups to the nascent unmethylated cytosines that reside opposite to methylated sites on the parental strand and thus the pattern of methylation is replicated. Right, demethylation: when a methyl group (M) is removed from methylated CpG sites.*

1.4 MAMMALIAN DNA MTases

The first mammalian DNA MTase reported was the Dnmt1. The *Dnmt1* cDNA from mouse has been cloned and its nucleotide sequence determined (Bestor, 1988; Tucker et al., 1996; Yoder et al., 1996). Partial loss of function mutation of the *Dnmt1* gene results in embryonic lethality showing that *Dnmt1* is essential for mouse development (Li et al., 1992). Dnmt1 contains a C-terminal (570 amino acids) domain joined to a much larger N-terminal (1051 amino acids) domain by a region of 13 alternating Gly and Lys residues; this region is encoded in the cDNA by 39 consecutive purine residues. This hydrophilic sequence is likely to form a flexible link between the N- and C-terminal domains. The C-terminal domain contains ten motifs (I-X) which are also present in most prokaryotes 5mC MTases. Motif I (PheXGlyXGly) was proposed to be part of the cofactor (SAM) binding site. This assignment was based on the presence of this consensus sequence in a wide variety of SAM dependent MTases, including N6-adenine, N4-cytosine, RNA and protein MTases (Ingrosso et al., 1989; Klimasauskas et al., 1989). Motif IV contains a dipeptide (ProCys) known to take part in the catalytic reaction. In the bacterial MTase HhaI this reaction involves the formation of a covalent intermediate between the cysteine from motif IV and the target cytosine (see section 1.1.2). The region between motifs VIII (GlnXArgXArg) and IX (ArgGlu) contains the target recognition domain (TRD) which determines the DNA sequence specificity (Klimasauskas et al., 1991) as well as the base to be methylated within the target sequence (Mi and Roberts, 1992). Several functions have also been localized to the N-terminal domain. The N-terminal domain contains a cysteine-rich region that is similar to metal-binding sites found in several regulatory proteins and it was shown to bind Zn (Bestor, 1992). The location of some of the relevant features in Dnmt1 are shown in Fig.1.7.

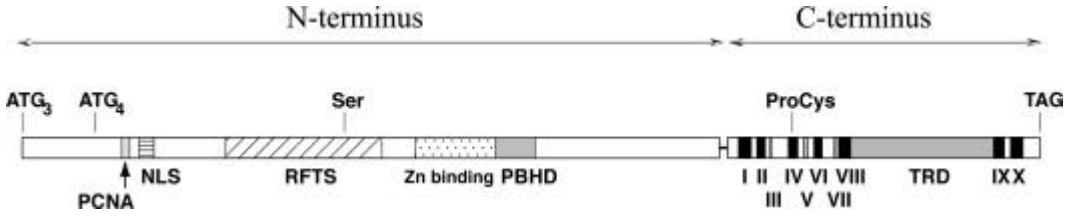


Fig.1.7. Structure of the mouse Dnmt1 protein*. The Dnmt1 protein contains an N-terminal domain of 1051 amino-acids which is joined to the 570 amino-acid C-terminal domain by a run of 13 alternating lysyl and glycyl residues, shown in the figure as a thick connecting line. Binding site for PCNA: proliferating cell nuclear antigen* (Chuang et al., 1997)*; NLS: nuclear localization sequence* (Cardoso and Leonhardt, 1999; Leonhardt et al., 1992)*; Ser: phosphorylation site at serine 514* (Glickman et al., 1997)*; RFTS: replication foci targeting sequence* (Leonhardt et al., 1992)*; Zn binding: in vitro zinc binding (Cys-rich) region* (Bestor, 1992)*; PBHD: polybromo-1 protein homologous domain* (Liu et al., 1998)*; ProCys: dipeptide part of the catalytic site* (Cheng et al., 1993; Wu and Santi, 1987)*; TRD: target recognition domain* (Trautner et al., 1988)*; filled boxes: highly conserved sequence motifs (I, II, IV, V, VI, VII, IX, X); grey boxes: less conserved sequence motifs (III, V, VII) between the mammalian and the bacterial MTases.*

A major phosphorylation site of the Dnmt1 has been reported (Glickman et al., 1997), indicating that the murine Dnmt1 is posttranslationally modified. The site lies in a region of the protein that is required for targeting of the MTase to the replication foci during S phase of the cell cycle (Leonhardt et al., 1992). Furthermore, it has been proposed that several factors might be involved in maintaining the methylation patterns in the

mammalian genome by interacting with Dnmt1. It has been reported that Dnmt1 binds to the proliferating cell nuclear antigen (PCNA) *in vitro* (Chuang et al., 1997) and this could mediate the association of Dnmt1 with the DNA replication machinery *in vivo.*

Proteolytic cleavage between the N- and C-terminal domains stimulates the *de novo* methylation activity of the murine Dnmt1 (Bestor, 1992) *in vitro*, suggesting that this enzyme may also be responsible for *de novo* methylation occurring *in vivo*. This was proposed to be the consequence of separation of the catalytic domain from the inhibitory N-terminal domain. This complex structure of Dnmt1 raises the possibility that Dnmt1 might carry out both, maintenance and *de novo,* methylation and these activities might be controlled by the expression of alternative isoforms or by posttranslational modification of the enzyme. Another alternative but not mutually exclusive possibility is that an independently encoded and developmentally regulated DNA MTase is responsible for *de novo* methylation. A null allele of the *Dnmt1* gene (Lei et al., 1996) showed that embryonic stem (ES) cells homozygous for the null mutation still had very low levels of genomic DNA methylation. Infection of mutant ES cells with the Moloney murine leukaemia virus (MoMuLV) resulted in *de novo* methylation of the integrated provirus DNA with a similar rate to that seen in wild type ES cells, suggesting that *de novo* methylation is not totally impaired by loss of Dnmt1. These results provide evidence that an independently encoded DNA MTase is expressed in early embryonic cells. Recently three additional mammalian MTases genes were cloned, *Dnmt2* (Okano et al., 1998; Van den Wyngaert et al., 1998; Yoder and Bestor, 1998), *Dnmt3a* and *Dnmt3b* (Okano et al., 1998). *Dnmt3a* and *Dnmt3b* are essential for *de novo* methylation and for mouse development and *Dnmt3b* is required for methylation of centromeric minor satellite repeats (Okano et al., 1999). However, the expression of *Dnmt3a* and *Dnmt3b* has not yet been characterized.

Evidence for the presence in some tissues of alternative Dnmt1 isoforms first came from the identification of a transcript longer than the one present in somatic cells during spermatogenesis (Trasler et al., 1992). This longer transcript could be due to additional sequences at the 5' end of the *Dnmt1* gene or to new alternatively spliced *Dnmt1* isoforms. Thus, different *Dnmt1* isoforms with different enzymatic properties might be expressed in specific tissues at particular stages and might play a role in the regulation of methylation during gametogenesis and tissue differentiation. The focus of this study was, therefore, to search for alternative *Dnmt1* isoforms and to study their expression during development, as well as to try to elucidate the Dnmt1 exon-intron boundaries and its relation with the protein structure.

The purpose of this section is to describe the principles and practical details of the techniques used in this study. Specific details of the methods used are described in the respective results section.

2.1 RNA ISOLATION

Total RNA extracted from mouse organs (strain C57Bl6) and cultured muscle cells (C2C12) was used as a template in RT reactions. The animals were killed by cervical dislocation and the organs were immediately taken, rinsed briefly in PBS and either frozen in liquid nitrogen or in isopentane precooled in liquid nitrogen depending on the subsequent applications, respectively RT or in situ hybridization.

During the RNA work a series of precautions were taken to avoid RNase contamination. Glassware was baked at 200^{o}C for at least 4 hours and water was treated with DMPC (Dimethylpyrocarbonate). A 10% DMPC stock was prepared in ethanol, which was then diluted 1:100 in double distilled water, left at room temperature overnight and finally autoclaved.

Total RNA extractions were done with a Qiagen kit (RNeasy) following the instructions of the manufacturer. The purity of total RNA was checked by running ~ 1µl of total RNA on a 1% agarose gel made in TAE buffer (0.04 M Tris-acetate/ 0.001 M EDTA pH 8) with 0.5 µg/ml ethidium bromide. The two ribosomal RNA bands (23S and 18S) were clearly detected when the gel was viewed under ultraviolet light (UV) and smears (as indicator of RNA degradation) were not seen between the two bands. To quantify the total RNA an aliquot was measured by UV absorbance at 260 nm (A_{260}) and 280 nm (A_{280}) where the absorbance of 1 in a 1 cm path length corresponds to a RNA concentration of 40 µg/ml. The absorbance ratio of 260 nm and 280 nm gave an estimate of the purity of the solution. Pure RNA solutions had A_{260}/A_{280} values between 1.7-2.

2.2 REVERSE TRANSCRIPTION

To assay for *Dnmt1* gene expression, total RNA from mouse tissues and cells was reverse transcribed into cDNA. For this reaction hexamer primers of random sequence as well as oligonucleotides of defined sequence were used. Random hexamers are used when a particular mRNA is difficult to copy in its entirety, because of the presence of sequences that cause the RT to stop synthesis. With this method, all RNAs in a population serve as templates for first-strand cDNA synthesis. Most specific is to use an oligonucleotide containing sequence information that is complementary to the desired target RNA. The advantage of using a specific primer is that only the desired cDNA is produced, facilitating the subsequent PCR amplification. The following components were mixed in a total volume of 20 µl to perform a reverse transcription reaction: 1 µg of total RNA was mixed with 1 x reaction buffer (100 mM Tris/ 500 mM KCl pH 8.3), 5 mM $MgCl_2$, 2-10 pmol of primer, 1 mM each dNTP and 20 units of avian myeloblastosis virus (AMV) reverse transcriptase enzyme (Roche Molecular Biochemicals). The mixture was then briefly vortexed and incubated at 25^{o}C for 10 min to allow the primer to anneal to the RNA, followed by incubation at 42^{o}C for 1 h during which the RNA was reverse transcribed resulting in cDNA synthesis. The AMV enzyme was then denatured by incubating the reaction at 99^{o}C for 5 min and then the mixture was cooled down to 4^{o}C for 5 min.

2.3 POLYMERASE CHAIN REACTION

PCR was used to amplify segments of DNA that were between two regions of known sequence using either genomic or cDNA as a template. Genomic DNA was prepared from mouse embryos as follows: i) the sample of interest was cut into small pieces using a blade and making sure that the sample stayed frozen at all times by keeping it in liquid nitrogen; ii) the pieces of sample were placed in an eppendorf tube and 850 µl of buffer (10 mM Tris pH 8.5/ 5 mM EDTA pH 8/ 0.2% SDS/ 200 mM NaCl) were added containing proteinase K at a final concentration of 150 µg/ml; iii) the tubes were placed at 50°C for 5 h or at 37°C overnight in a shaking water bath and from time to time the samples were vortexed to disrupt the tissue; iv) the samples were then heated at 95°C for 10 min.

For the PCR reactions , either 5 µl from the RT product (~0.25 µg of cDNA) or 100 ng genomic DNA were used in a total volume of 50 µl containing 1 x PCR buffer (100 mM Tris/ 50 mM $MgCl_2$/ 500 mM KCl pH 8.3), 0.2 mM each dNTP, 15 pmol from each primer and 5 units of Taq polymerase. Different Taq enzymes were used depending on the size of the fragment expected: for up to 2000 bp, the Taq polymerase was used; for bigger fragments, the Long Expand PCR kit (Roche Molecular Biochemicals), which contains a mixture of a thermostable Taq and Pwo DNA polymerases with a 3' to 5' proofreading activity, was used. For the amplification of cDNA, oligos located in cDNA areas spanning intron locations were used to control for possible genomic DNA contamination.

The PCR mixtures containing all the necessary reagents were denatured by heating at 94°C, then cooled to a temperature which allows the oligonucleotide primers to anneal to their target sequences (usually 5°C less than the primer's melting temperature), finally the annealed primers were extended with DNA polymerase. The cycle of denaturation, annealing and DNA synthesis was repeated for ~ 30 cycles, so that the products of one round of amplification served as templates for the next, doubling the amount of desired DNA product in each round. The reactions were carried out on a DNA Thermal Cycler (Biometra) and when optimization of annealing temperatures was necessary the Mastercycler Gradient (Eppendorf) was used. The latter allows the use of a gradient function which enables changes of temperature between 4°C and 99°C to be tested simultaneously.

The amplified DNA fragment was separated by gel electrophoresis on a 1% agarose gel in TAE buffer containing 0.5 µg/ml ethidium bromide. The fluorescence from the DNA bands with the intercalating dye was visualized by exposing the gel to UV light. Depending upon the expected size, the following DNA molecular weight markers were used: SmartLadder (Eurogentec), 100 bp Ladder, Lambda DNA cut with BstEII or Lambda DNA cut with Hind III (all from New England BioLabs). Comparison of the fluorescence intensity of the PCR fragments with the standard bands from the molecular weight marker provided, in addition, an estimation of the amount of DNA in the samples.

Fresh PCR products of the desired size were gel purified with Qiaex II (Qiagen). The gel purification method is based in the adsorption of DNA molecules to silica particles in the presence of high salt. After washing away the non-nucleic acid impurities, the DNA was eluted in ~ 20 µl of 10 mM Tris-HCl pH 8.5. Fresh PCR products were also directly ligated into the appropriate vector depending on the subsequent applications (pCRII vector for sequencing or *in vitro* transcription, pCR3.1 vector for protein expression in mammalian cells, both vectors are from Invitrogen).

2.4 PLASMID DNA TRANSFORMATION INTO *ESCHERICHIA COLI* AND SELECTION OF RECOMBINANTS

The plasmid DNA was transferred into bacteria by a process known as transformation, in which the plasmid is taken up by bacteria that had been pretreated to make some of the cells temporary permeable to small DNA molecules, i.e., competent (Dagert and Ehrlich, 1979; Mandel and Higa, 1970). The new phenotype conferred by the plasmid (i.e., resistance to an antibiotic) allows selection of bacteria that had been successfully transformed. After transformation, the *Escherichia coli* cells were plated on Luria Bertani (LB) medium with the appropiate antibiotic and incubated overnight at 37^{o}C. The next day, isolated bacterial colonies were checked either by PCR using oligos located in the vector and insert or by small scale plasmid DNA isolation using the alkaline lysis method (Birnboim and Doly, 1979) and subsequent restriction enzyme analysis. Briefly, bacteria were lysed by treatment with a solution containing SDS (denatures bacterial proteins) and NaOH (denatures chromosomal and plasmid DNA). The mixture was neutralized with potassium acetate, causing the plasmid DNA to reanneal rapidly. Most of the chromosomal DNA and bacterial proteins precipitated, as well as the SDS which formed a complex with potassium, and were removed by centrifugation. The reannealed plasmid DNA was then concentrated by ethanol precipitation. Restriction enzyme and/or PCR analyses of the plasmid DNA allowed identification of the plasmids containing the right insert, i.e., of the correct recombinants. Selected recombinant bacteria were grown in large scale and subsequent isolation of plasmid DNA was carried out with the Plasmid Maxi kit following the instructions of the manufacturer (Qiagen). Briefly, this method is based on a modified alkaline lysis procedure, followed by binding of plasmid DNA to an anion-exchange resin under appropriate low salt and pH conditions. RNA, proteins and low molecular weight impurities were removed by a medium salt wash. Plasmid DNA was eluted in a high salt buffer, concentrated and desalted by isopropanol precipitation.

2.5 SEQUENCING

Plasmids were sequenced to verify that they contained the desired inserts and to control for unwanted mutations. Samples were processed by the dye terminator method (Applied Biosystems) and purified from unintercalated labeled dNTPs by Centri-Sep spin columns (Applied Biosystems). The samples were then mixed with 4 µl buffer (deionized formamide and 25 mM EDTA pH 8 containing 50 mg/ml Blue dextran in a ratio 5:1 formamide:EDTA-Blue dextran), denatured by heating at 94^{o}C for 4 min, and loaded onto a sequencing gel on an automated DNA sequencer (Applied Biosystems). This sequencing method is based on the dideoxy method developed by Sanger (Sanger et al., 1977). Briefly, this technique involves the synthesis of a DNA strand from a single-stranded template by DNA polymerase and depends on the fact that dideoxynucleotides (ddNTPs) are incorporated into the growing strand in the same way as the conventional deoxynucleotides (dNTPs). However, ddNTPs differ from dNTPs because they lack the 3'-OH group necessary for chain elongation. With the dye terminator only molecules that have terminated are labeled, and the four base-specific terminations are carried out in the same tube and then analyzed in the same lane of a gel. The latter is possible because a different fluorescent dye is attached to each nucleotide and incorporated into the DNA chain during the copying reaction catalyzed by DNA polymerase. The samples were then separated on a polyacrylamide gel and the fluorescent nucleotides excited with a laser beam. The emitted fluorescence was collected by detectors and sent to a computer, where the appropriate software converted the data into nucleotide sequence. For

sequence assembling, editing and alignments, the Lasergene software programs Edit Seq, SeqMan and MegAlign (DNASTAR, Madison) were used. Some samples were sequenced by the companies GATC or MWG.

2.6 NORTHERN HYBRIDIZATION

This technique was used to study the number, size and levels of *Dnmt1* transcripts in different tissues and it is based on the hybridization of a labeled nucleic acid probe (RNA or DNA) to a complementary sequence of mRNA.

2.6.1 PROBE LABELING

cDNA fragments subcloned into pCRII vectors were prepared by restriction digestion of plasmids and isolated by gel purification. For radioactive labeling of DNA fragments the random primer extension method (Amersham) was used and the enzyme required for this process was *Escherichia coli* DNA polymerase I (Klenow fragment). The procedure relies on the ability of random oligonucleotides to anneal at multiple sites along the length of a DNA template. Starting from these oligonucleotides annealed to the DNA, the DNA polymerase I synthesizes new complementary strands incorporating labeled dNTPs present in the reaction mixture.

The reaction mixture contained: 25 ng of DNA (which have been previously gel purified) and 5 µl of random nonamer primers (Amersham), which were brought to 50 µl final reaction volume with double distilled water. This initial mixture was denatured by boiling for 5 min. Keeping the tube at room temperature, the unlabeled dNTPs (0.1 mM each, omitting those to be used as label) and reaction buffer (from Amersham, containing Tris-HCl pH 7.5, 2-mercaptoethanol and $MgCl_2$) followed by the radiolabeled dNTPs and DNA polymerase I (Klenow fragment) enzyme (0.04 units/µl) were added. The reaction was then incubated at 37°C for 1 h. Phosphorus-labeled dNTPs (3000 Ci/mmol [α–^{32}P] dCTP and 3000 Ci/mmol [α–^{32}P] dATP) were used. Non incorporated labeled nucleotides were removed from the probe by precipitation with 0.1 volume of 4 M LiCl and 2.5 volumes of chilled absolute ethanol and incubation at –70°C for 30 min and centrifugation at 13000 g for 15 min at 4°C. The pellet was then washed with ~ 100 µl of cold 70% ethanol, air dried and dissolved in 20 µl of double distilled water. The efficiency of the labeling reaction was monitored by a scintillation counter (Beckman).

2.6.2 HYBRIDIZATION

Hybridization and washing conditions varied depending on the stringency required to obtain the highest signal and the lowest background. The hybridization solution is one of the critical parameters to consider and should contain blocking reagents such as herring sperm DNA, as blocking reagent, to prevent nonspecific binding of RNA probes to the blotting membrane. The hybridization temperature also needs to be optimized to ensure optimal hybridization to the target and minimal nonspecific binding of the probe. The final, stringent wash is usually 10°C below or in some cases the same as the hybridization temperature. The buffer used was from Clontech or according to Church *et al*. (Church and Gilbert, 1984) (7% SDS/ 1% BSA/ 0.25 M $NaHPO_4$ pH 7.2/ 1 mM EDTA pH 8). Prehybridization which contained only the hybridization buffer (5 ml) without the labeled

probe was done for 4 h and hybridization overnight, containing 5 ml hybridization buffer and 2 x 10^6 cpm/ml labeled probe. The probe and 50 µg/ ml of herring sperm DNA were denatured by boiling for 5 min before adding to the blots. The temperature for pre- and hybridization was 68^{o}C. After hybridization the blots were washed twice 20 min with 0.1x SSC, 0.1% SDS at 50^{o}C with continuous shaking. The blot was retrieved from the hybridization bottles and the excess of wash solution was removed not allowing the membrane to dry. The blot was covered with plastic wrap and exposed to a phosphorimager screen (Molecular Dynamics) for 24 h to one week. The signals were detected by scanning the screens on a phosphorimager using the Image Quant software (Molecular Dynamics). The probe was removed from the blot by incubating for 10 min in boiling distilled water. This process was repeated three times and the water allowed to cool for 10 min before the blot was stored in plastic wrap at –20^{o}C, until needed for new hybridization experiments.

2.7 IN SITU HYBRIDIZATION

The principle of in situ hybridization is very similar to Northern hybridization and is based on the annealing of a labeled nucleic acid probe (RNA or DNA) to a complementary sequence of mRNA. The two techniques differ in that the starting material for a Northern is a tissue extract, whereas the primary material for in situ hybridization is a histological tissue section. The relative cell position in the tissue is lost with Northern blots and mRNA levels are averaged from all the cells contained in the original sample. However, in situ hybridization can detect amounts of mRNA contained in a single cell. Furthermore, since in situ hybridization is a histological technique, tissue structure is preserved and it is possible to precisely identify the cell type expressing the gene of interest. A major advantage of in situ hybridization is that it allows the maximal use of tissues that may be in short supply while an organ extract might yield sufficient RNA for one or two Northern blots only. These Northern blots are useful to quantify specific mRNAs, but large amounts of tissue extracts are required. Although Northern blots can be probed multiple times for different mRNAs, in fact this is limited to four or five different probes at most, since repeated stripping of the blots leads to RNA loss from the membranes. In situ hybridization is a semi-quantitative technique but provides information concerning the cellular source of the RNA.

There are many different ways to do in situ hybridization and it can be done with radioactive or non-radioactive probes. The use of ^{35}S labeled riboprobes to hybridize frozen tissue sections has been proven to be the most sensitive method (Wilcox, 1993). ^{32}P labeled probes used for Northern blots are not suitable for the in situ technique due to the high energy produced by ^{32}P, which does not allow the detailed cellular resolution of the signal. To control for genomic DNA hybridization and unspecific binding, concurrently with the slides containing the antisense probe, adjacent tissue sections were also included containing a control sense RNA probe.

2.7.1 PROBE LABELLING

For riboprobe preparation 1 µg of plasmid DNA containing T7, SP6 or T3 RNA polymerase binding sites was linearized by digestion with a restriction enzyme downstream (antisense probe) or upstream of the insert sequence (control sense probe). The linearized DNA template was then purified by two phenol extractions (Phenol/Chloroform/Isoamylalcohol: 25/24/1) and LiCl/ethanol precipitation. The DNA was transcribed at 37^{o}C in 20 µl total volume for 2 h with 2 µl 1x transcription buffer

(Promega) (200 mM Tris-HCl pH 7.5/ 30 mM $MgCl_2$/ 10 mM spermidine/ 50mM NaCl), 2 µl 10mM GTP/ATP, 2.5 µl each α-^{35}S CTP (800 Ci/mmol), α-^{35}S UTP (800 Ci/mmol) and 1 µl of the appropriate RNA polymerase enzyme (20 units). The labeled probe was recovered by LiCl/ethanol precipitation and washed with ethanol as described in preparation of probes for Northern hybridization (section 2.6.1). A sense probe, was also labeled as a negative control.

Long probes penetrate less effectively into fixed tissue and the magnitude of this effect depends on the tissue preparation and type of probe used. To increase penetration of the probe in the tissue, long fragments used as RNA probes (~ 1000 nucleotides) were reduced to an average of 150 to 200 nucleotides by alkaline hydrolysis. The alkaline hydrolysis was carried out by dissolving the probe in 50 µl of DMPC treated water and the pH adjusted to 10.2 by addition of 30 µl of 0.2 M Na_2CO_3 and 20 µl of 0.2 M $NaHCO_3$ and incubated at 60^{o}C for 30 min. The hydrolysis was stopped by addition of 3 µl of 3 M sodium acetate pH 6 and 5 µl of 10% glacial acetic acid. The hydrolyzed probe was precipitated with ethanol and resuspended in a small volume of DMPC water.

2.7.2 SAMPLE PREPARATION

Tissue sectioning

Handling of glass, reagents and the extraction of organs from the animal was as in section 2.1. If the tissue was dense and compact (e.g. muscle), the specimen was frozen in isopentane precooled in liquid nitrogen. Isopentane shortened the time needed for freezing. Fragile tissues were placed in a plastic capsule containing a freezing agent (Tissue Tek), before being frozen. Frozen tissues were sectioned with a cryotome (Microm) into 6-12 µm thick slices, thaw-mounted at 42^{o}C on coated slides (Superfrost/Plus, Fisher Scientific) and immediately re-frozen by placing the slides in the plastic slide boxes kept in the cryostat or on dry-ice and stored at –70^{o}C until use. The frozen tissue slides were first left at room temperature for 1 h to thaw before tissue sectioning.

Fixation

Unlike blotting membranes, tissues are not chemically uniform and respond differently to fixation protocols. Procedures that preserve good morphology and retain RNA usually do not give high hybridization efficiency. Extensive fixation preserves histological detail and prevents loss of target RNA during hybridization and wash procedures. However, milder fixation increases accessibility of the hybridization probes to the RNA target (and to any other molecules such as antibodies) and reduces the possibility of chemical modification of the target RNA and concomitant loss of hybridization. Thus it is important to optimize fixation and prehybridization treatments of tissue to find the best compromise between these requirements. Cross-linking fixatives were favored because they provided much better retention of cellular RNA and help to inactivate nucleases in tissues.

The frozen tissue sections were fixed for 20 min in 4% paraformaldehyde (8 g of paraformaldehyde were dissolved in 200 ml PBS by adding 1-2 pellets of NaOH, the pH was then adjusted to 7 with HCl). The sections were then washed three times 3 min in PBS.

Acetylation

The purpose of this step was to cover the section with negative charges to repel nucleic acids and thus reduce electrostatic binding of the probe to the section. The tissue slides were incubated for 10 min in the acetic anhydride solution (250 ml distilled water/ 3.1 ml Triethanolamine/ 675 µl acetic anhydride). After the incubation period the slides were washed 3 min in PBS.

Dehydration

The sections were dehydrated by passing the slides through an ascending ethanol series (50%, 70%, 90%, 96%) and were then left to dry at room temperature for 2-3 h.

2.7.3 HYBRIDIZATION

The slides were placed in a box with a piece of filter paper saturated with buffer (4 x PBS/50% formamide) during the prehybridization and hybridization steps. The prehybridization step was carried out by covering each slide with 100 µl of hybridization buffer (50% deionized formamide/ 10 mM Tris-HCl pH 7.5/ 5 mM EDTA pH 8/ 2 x SSC/ 150 µg/ml herring sperm DNA/ 150 µg/ml yeast tRNA/ 10% dextran sulfate/ 1 x Denhardt´s solution/ 0.1 mM UTP/0.1 mM CTP) without probe and incubating at 37°C for 4 h. Hybridization was performed with a mixture containing the hybridization buffer and the ^{35}S-labeled riboprobe (10^8 cpm/µl). This mixture was heated at 95°C for 10 min and placed on ice. DTT was then added at a final concentration of 0.1 M. The probe mixture was vortexed and 10 µl was added to the sections. Siliconized coverslips were placed on top of each sample and incubation was done overnight at 42°C. The coverslips were removed from the sections by soaking the slides in 1x SSC.

After hybridization the sections were washed 20 min at 60°C in 5 x SSC/ 1 mM DTT, 30 min at 60 °C in 2 x SSC / 50% formamide/ 1mM DTT, 2 x 10 min at 37°C in RNase-buffer (10 mM Tris-HCl pH 8/ 1mM EDTA pH 8/ 0.5 M NaCl), 30 min at 37°C RNaseA digest (10 mg/200 ml) to remove nonspecifically bound single-stranded RNA probe, thus reducing the background signal in autoradiographs , 10 min at 37°C in RNase buffer, 1h at 60°C in 2 x SSC, 10 min at 60°C in 0.1 x SSC.

The slides were dehydrated by dipping them for a few minutes in an ascending ethanol series and were left to dry at room temperature. The slides were exposed to a Phosphorimager screen to produce a fast low resolution image or dipped in photographic emulsion for high cellular resolution images. The latter involved dipping of slides into a nuclear track emulsion (Amersham), whereby the sections are covered with a thin layer of the emulsion (silver bromide crystals in gelatin). The emulsion was first melted at 42°C and after dipping the slides in it, these were left to dry at room temperature for 1 h and then exposed in the dark at 4°C. Exposure times varied from 1 to 4 weeks depending on the probe and sample. To avoid artifacts, the slides were allowed to warm up to room temperature before photographic development. Decay of radionucleotides leads to the formation of latent images in the emulsion layer. Latent images were converted into real images by photographic development (Kodak) and fixation (Kodak). The slides were then counterstained with hematoxylin (Sigma).

2.8 CELL CULTURE AND TRANSFECTION

The cells used were mouse C2C12 myoblasts (Yaffe and Saxel, 1977) and COS-7 cells (African green monkey kidney fibroblast-like cells transformed with SV40T antigen; (Gluzman, 1981)). The C2C12 cells were used because they are a well characterized tissue culture skeletal muscle differentiation system to screen for *Dnmt1* isoforms during myogenesis. These cells were studied at different stages of growth and differentiation (see Results section). The COS-7 cells were used to test the translation efficiency of *Dnmt1* isoforms by immunofluorescence and Western blot analyses (see Results section).

Both types of cells were grown in a humidified atmosphere of 5% CO_2 and at a temperature of 37^{o}C. The cells were maintained at subconfluent densities in growth medium containing Dulbecco´s modified Eagle´s medium (DMEM) with 10% and 20% fetal bovine serum (FBS) for COS and C2C12 cells respectively, and were split every 48-72 h. Proteolytic enzymes such as trypsin in combination with EDTA were used to detach the cells from the growth surface to harvest or to subculture them. Frozen cells were thawed by placing the vials in a 37^{o}C water bath by gently agitating the vial, cell suspension was then pipetted into a plate containing pre-warmed growth medium. For long term storage cells were harvested as above and washed once with complete medium. The cells were then resuspended in complete medium and counted. The cells were stored in complete medium with 10% DMSO and aliquots of cells (about 10^{6} cells/ml/cryovial) were frozen and kept in liquid nitrogen. C2C12 cells were transferred to DMEM with 5% horse serum (differentiation media) whenever differentiation was needed.

Transient transfections were done with the corresponding eukaryotic expression plasmid in COS-7 cells by the lipofection method (GenePorter), following the instructions of the manufacturer (Gene Therapy Systems, Inc.). Seventy two hours after transfection, cells were washed twice with PBS and either fixed with 3.7% formaldehyde in phosphate buffered saline and immunostained or pelleted for subsequent cell extract preparation and Western blot analysis.

2.9 IMMUNOFLUORESCENCE ANALYSIS

2.9.1 TISSUE SECTIONS

Organs were collected from the animal as in section 2.1 and fixed overnight with Serra's (60% absolute ethanol/ 30% formaldehyde/ 10% glacial acetic acid) or Carnoy's solution (60% absolute ethanol/ 30% chloroform/ 10% glacial acetic acid). Both have proven to give equally good results. The next day the organs were embedded in paraffin with a Techno-Tec1 automat (Pathotec).

The sectioning of paraffin embedded material was done with a rotary microtome (Microm). After each cut the section was floated on a thermostatically controlled water bath (40-45^{o}C). Then a slide was dipped under the section and upon lifting the section adhered to the slide. The slides were then left to dry on a warming plate.

Sections were deparaffinized by incubating at 60^{o}C for 1 hour, then in Xylol for 15 min, in isopropanol for 5 min, in a descendent ethanol series (96%, 70%, 50%) 5 min each and finally in PBS for 5 min. The samples were then placed in 0.01 M citrate buffer pH 6 and autoclaved for 20 min at 94^{o}C.

2.9.2 CULTURED CELLS

Cells were treated were treated with 3.7% formaldehyde in PBS for 10 min which leads to the formation of chemical cross links between free amino groups. Fixation with formaldehyde does not allow access of the antibody to the specimen and therefore was followed by a permeabilization step using a nonionic detergent. After fixation the coverslips were washed twice with PBS. Permeabilization of the fixed cells was performed by incubation in 0.2% Triton X100 in PBS for 5 min at room temperature. The cells were finally rinsed and kept in PBS.

2.9.3 IMMUNOSTAINING

Fixed samples (tissues or cells) were blocked for 30 min with 0.2% fish skin gelatin. After fixation, incubation with the first antibody (Ab) for 2 hours was done. The unbound antibody was removed by three 10 min washes with PBS-Tween-20 (0.2%). The secondary fluorescently conjugated Ab was then added and incubated for 1 h. After washing as before, nuclear DNA was stained for 5 min with Hoechst 33258 at a concentration of 1μg/ml. Finally, samples were washed again and mounted in Mowiol, which hardens within a few hours forming a permanent preparation.

Secondary antibodies with different fluorochromes (fluorescein or rhodamine) were used. Under the appropriate illumination, fluorochromes are excitated alllowing the localization of the antibodies and, through them, the distribution of the antigen under study. The methods used are indirect detection and relied on a labeled secondary antibody that recognizes specifically the unlabeled primary antibody.

Protein expression and subcellular localization was also analyzed by epitope tagging. The tag itself is a small open reading frame, and fusion of tags with the protein under study was achieved by standard molecular biology methods. The epitope tag is a short amino acid sequence, that is placed within the coding region of the target protein to specify the binding site for a known monoclonal or polyclonal antibody. The tagged target protein can then easily be detected using well characterized antibodies directed to the tag. Tagged proteins were then produced in COS-7 cells using an appropriate expression vector (pCR3.1 from Invitrogen). The expression and localization of tagged proteins were monitored in transfected COS-7 cells. The Flag tag is not derived from any known protein and contains an 8-amino acid peptide (Asp, Tyr, Lys, Asp, Asp, Asp, Asp, Lys) that is hydrophilic and can be placed at both the amino and carboxyl termini of the target protein. There are two commercially available antibodies, M1 and M2, that recognize the Flag tag. The M1 antibody requires the amino terminus of the epitope to be available, thus the tag is placed at the amino terminus of the protein. The M2 antibody was the one used, which recognizes the same Flag epitope when placed internally within a protein or at the carboxyl terminus.

2.10 MICROSCOPY

Immunofluorescence samples were examined and photographed on a Zeiss Axioplan 2 microscope equipped with phase-contrast and epifluorescence optics, using 63x Plan-Apochromat, 2.5x, 20x, 40x Plan-Neofluar and 4x Achroplan objectives.

A Zeiss AxioLab microscope equipped with phase-contrast and darkfield condenser, was used to examine in situ hybridization specimens using 4x, 10x, 20x and 40x Achroplan

and 2.5x Plan-Neofluar objectives.

Images were taken with CCD cameras (SensiCam, PCO and ProgRes, Kontron Elektronik) using Zeiss Axiovision and Photoshop software.

2.11 CELL EXTRACTS AND WESTERN BLOT ANALYSIS

Transfected cells were harvested by centrifugation at 13,000 x rpm for 5 min and cell pellets were resuspended in ice-cold RIPA buffer (500 mM Tris-HCl pH 8/ 150 mM NaCl/ 1% NP-40/ 0.5% DOC/ 0.1% SDS) with protease inhibitors (leupeptin, aprotinin, pepstatin). The cell suspension was incubated on ice for 5 min and extracted proteins were recovered from the supernatant after centrifugation at 13000 x rpm for 5 min. The protein extracts were denatured by boiling in Laemmli sample buffer (2% SDS/ 20% glycerol/ 250 mM Tris-HCl pH 6.8/ 10% β-mercaptoethanol/ 0.1% bromophenol blue) prior to loading onto a gel. Proteins were separated on a 10-20% gradient SDS polyacrylamide gel and transferred to a polyvinylidene fluoride (PVDF) membrane (BIO-RAD) with transfer buffer (48 mM Tris/ 39 mM glycine/ 0.037% SDS and 20% absolute methanol) using a semidry blotter (Hoefer, 1h at 60 mA). The PVDF membrane was blocked by treatment with 5% nonfat milk powder in PBS (blocking buffer) for 30 min and then incubated in a solution of the primary antibody diluted with blocking buffer and 0.2% Tween 20 for 2 h at room temperature. The blot was washed in PBS containing 0.2% Tween 20 three times 10 min, incubated with the secondary antibody for 1 h and finally with Streptavidin-HRP for half an hour. The signals were detected with the ECL+ reagent according to the instructions of the manufacturer (Amersham). Primary and secondary antibodies were used at the following dilutions: 1:5,000 anti-FLAG antibody M2 (Kodak); 1:10,000 anti-mouse IgG biotin (Amersham); 1:500 Streptavidin-HRP (Amersham). The signals were recorded on a luminescent image reader (Fuji). Images were assembled and annotated using Adobe Photoshop and Adobe Illustrator on a Power Macintosh computer.

3.1 GENOMIC STRUCTURE OF *DNMT1* GENE

The Dnmt1 protein has a C-terminal (~500 amino acids) domain, sharing motifs (I-X) with over 100 prokaryotic 5mC MTases, linked to a N-terminal (~1000 amino acids) regulatory domain (see Fig. 1.7). The sequence similarities among mammalian and bacterial DNA MTases suggest a common evolutionary origin. A decade ago, it was proposed that the mammalian enzyme evolved via fusion of an ancestral prokaryotic restriction MTase gene with a second gene based on cDNA sequence and on similarity of the putative peptide to bacterial MTases (Bestor, 1990). However, based on comparison of several eukaryotic MTases and deletion mapping of the enzymatic activity, we have suggested that MTases evolved by fusion of at least three separate genes (Margot et al., 2000).

At the beginning of this work additional 5' sequences were identified in the *Dnmt1* cDNA (Tucker et al., 1996) as well as two different mRNA transcripts (Trasler et al., 1992): one present only in testis (pachytene spermatocyte) and a second one characteristic of all somatic tissues.

In order to relate the functional motifs in the Dnmt1 protein with the exon-intron boundaries of the *Dnmt1* gene and to study the regulation of the new isoforms, the elucidation of the genomic structure of the mouse *Dnmt1* gene was undertaken. Furthermore, the sequence of the 5' end of the *Dnmt1* gene was determined to establish the basis for investigating promoter regulation of the different *Dnmt1* isoforms.

Overlapping phage clones, isolated from screening a mouse 129/Sv genomic library and subcloned into pSP72 (Promega) or pBluescript KS (pMET in Fig. 3.1, Stratagene), were generously provided by Dr. En Li (Charlestown, USA) (Li et al., 1992). The region between pSP72RV10 and pSP72S5 was obtained by genomic PCR using DNA prepared from frozen 12 day old mouse embryos from strain C57BL6. The genomic DNA was extracted as indicated in the Materials and Methods section. One hundred ng of DNA in a reaction volume of 100 µl was amplified with a mix of Taq and Pwo DNA polymerases (Roche Molecular Biochemicals). The ~ 6 kb PCR product was directly cloned into pCRII vector (Invitrogen) and the resulting plasmid designated pCg6. The region upstream of pSP72S15 (pCg1 and pCg3) was obtained by a combination of genomic PCR and genome walking. The genome walking approach was used to obtain the fragment upstream the oocyte specific exon (pCg1) by using genomic DNA, which have been digested with five different enzymes and the resulting restriction fragments ligated to an adaptor primer. The upstream walk towards the putative promoter from the oocyte exon was then carried out following the instructions of the manufacturer (Clontech). PCR reactions with reverse primers located in the oocyte exon together with the adaptor primers of the genomic DNA fragments were done. The genomic PCR carried out to obtain the pCg3 fragment was done with a forward primer located on pCg1 and a reverse primer located on pSP72S15.

In order to determine the sizes of the introns, sequencing (GenBank accession numbers: AF 175410-175431, AF 234317-234318) and/or gel electrophoresis of the PCR products (Fig. 3.1) were performed. The PCR products containing genomic fragments were generated using primers located on both sides of the putative exon boundary. The mouse *Dnmt1* cDNA was completely resequenced and the revised sequence has been submitted to GenBank (accession number AF 162282). The determination of the exon-intron boundaries (Table 3.1) was carried out by comparing genomic and cDNA sequences. Intron localizations at the 3' end of the *Dnmt1* gene were obtained in collaboration with the group of Dr. R. J. Roberts (Beverly, USA).

In this thesis it has been elucidated that the mouse *Dnmt1* gene consists of 39 exons and spans over 56 kb (Fig. 3.1). The exon sizes vary between 32 (exon 7) and 352 bp (exon 39) (Table 3.1). The region between the exon 1 and exon 29 contains introns of widely

different sizes as well as two very large introns. This region has a low exon density (number of exons per unit length). However, the region between exon 30 and 39 harbors introns that are smaller and much more uniform in size (Fig. 3.1), this region clearly has a higher exon density. The two different exon density patterns correspond to the division between catalytic and regulatory domains of the protein proposed before (Bestor, 1990). Furthermore, the first exon is located in a CpG island (Fig. 3.8) and is separated from exon 2 by a 9.3 kb long intron, which is a typical feature of house keeping genes (Cross and Bird, 1995). The region spanning pCg1 until the end of exon 5 (Fig. 3.1) was completely sequenced (GenBank accession numbers AF 175410-175412). The TAG stop codon was found 39 bp into exon 39 and two polyadenylation signals (AATAAA) were found, the first one was mapped 290 bp downstream of the stop codon and the second one was identified 274 bp further downstream.

During the process of this work the structure of the human *Dnmt1* gene has also been published (Ramchandani et al., 1998). A comparison between mouse and human DNA sequences reveals a similar genomic organization. Exon boundaries are conserved in 35 out of 38 cases. The only exception to a perfectly conserved intron-exon organization is the presence of an additional intron in the human gene whose corresponding location would be in mouse exon 37. Intron sizes, with the exception of two large introns after oocyte and spermatocyte specific exons, range between 77 bp and ~ 3kbp (Table 3.1). All 5' and 3' splice sites agree with the consensus found for rodents (Senapathy et al., 1990; Shapiro and Senapathy, 1987) with the exception of the 5' donor site of intron 10 where the GT consensus is replaced by GC (Table 3.1). The human *Dnmt1* gene has two exceptions to the GT/AG rule (Table 3.1), but none correspond to the one in the mouse gene. In one case GC replaces GT as splice donor of intron 11, while in the other case CT replaces AG as splice acceptor of intron 39 (Ramchandani et al., 1998).

The location of the conserved motifs of the catalytic domain (Fig. 1.7) (Leonhardt and Bestor, 1993) was compared with the intron-exon boundaries (Table 3.1). Three out of seven exon boundaries fall within highly conserved motifs and thus indicate that there is no correlation between proteins motifs and exon-intron boundaries.

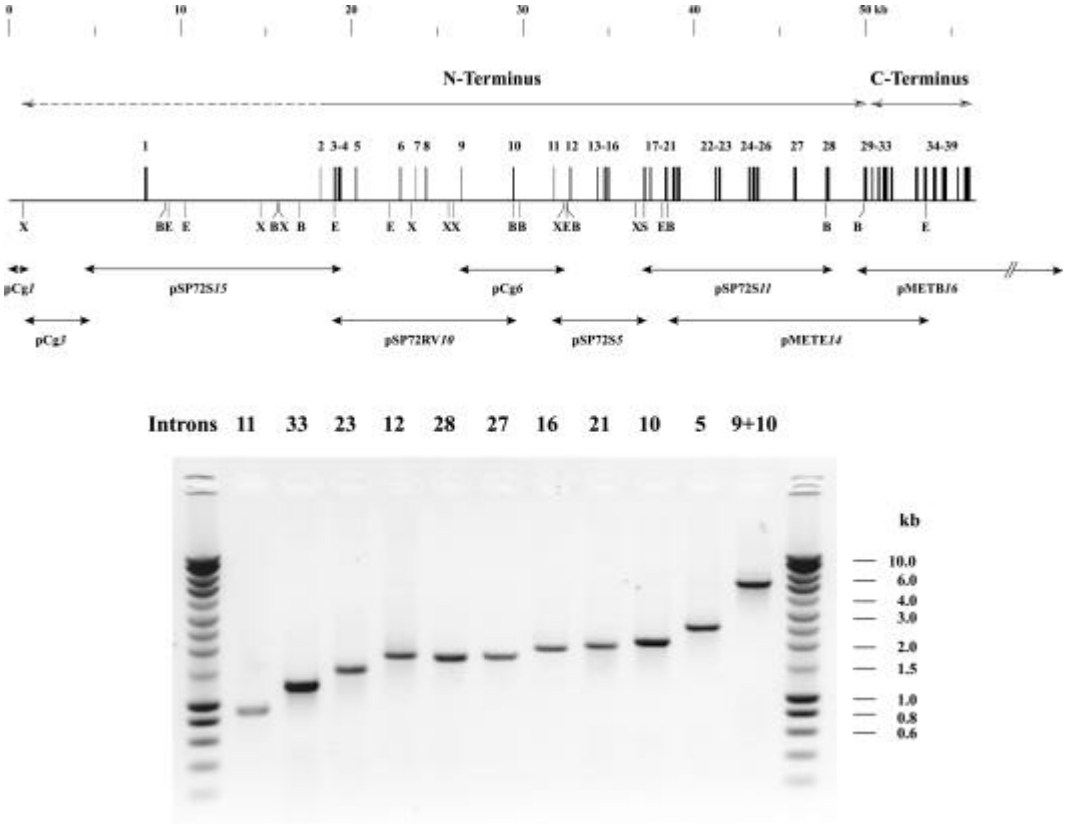


Fig. 3.1. Genomic organization of the mouse Dnmt1 gene. *At the top of the figure are depicted the exons and introns, the restriction map and the localization of the cloned genomic DNA fragments within the Dnmt1 locus. B: BamHI; E: EcoRI; S: SalI; X: XbaI. The clones denominated pSP72S5 and pSP72S11 share the SalI site in exon 17. Plasmid insert sizes are indicated in italics. At the bottom of the figure the intron size determinations by agarose gel electrophoresis are shown. MTase specific primers were selected within 50 bp on both sides of the putative intron boundary and used for PCR amplification. The PCR products were separated on an agarose gel and intron sizes were determined from the standard curve of the molecular weight markers. The results are summarized in the table 3.1.*

	EXON		INTRON			EXON
#	Size	5' Donor Site	Size	Phase	3' Acceptor Site	
						gccccc..
1	160	..cgc agg CG	g t a g g t..	10.098	2	..a c c c a g	G ctc aaa..
2	37	..gaa AAG	g t a a a c..	0.762	0	..c t a c a g	gag tgt..
3	108	..tct GAG	g t a a g t..	0.115	0	..t t g t a g	gaa ggc..
4	211	..acc ctt T	g t a a g g..	0.818	1	..t t t a a g	CA gtt gaa..
5	79	..acg aag GG	g t g a g t..	2.5	2	..t g a c a g	C ccc act..
6	76	..gac CAG	g t g g g c..	0.876	0	..g c c c a g	gat aag..
7	32	..aca gag AG	g t a a g a..	0.539	2	..t g a c a g	T ggt gct..
8	79	..gaa CAG	g t a a a g..	1.974	0	..t c c t a g	gaa gat..
9	41	..aga gag CT	g t a a g t..	3.2	2	..t g g c a g	A tca ttg..
10	85	..aaa AAG	g c a a a g..	2.1	0	..t g t c a g	gat aaa..
11	41	..aga gat CC	g t a a g t..	0.9	2	..t t g t a g	A gct gcc..
12	82	..gag AGG	g t a a g t..	1.8	0	..g t t t a g	gag gag..
13	65	..cag agc AG	g t g a t g..	0.258	2	..t t c c a g	A tcg gag..
14	43	..ccg AAG	g t a a g t..	0.094	0	..c c a c a g	atc aac..
15	81	..gat GCT	g t g a g t..	0.078	0	..c t t c a g	gtg gat..
16	110	..tcc ttc AG	g t a a t c..	1.9	2	..a t t c a g	T gtg tac..
17	119	..atg gaa G	g t a t g t..	0.228	1	..c t a c a g	GT ggt att..
18	93	..tcc act G	g t g a g t..	0.8	1	..t c c t a g	CA ttt gct..
19	152	..att GAG	g t g c g a..	0.284	0	..t t a a a g	acc act..
20	188	..gga cag AG	g t a a g g..	0.093	2	..c c t c a g	G cga gca..
21	178	..tgt GAG	g t g a g c..	2	0	..t t t c a g	gtc tgt..
22	98	..aag agg AG	g t a g g c..	0.121	2	..c t g c a g	G tgt cct..
23	148	..atg AAG	g t g a g g..	1.5	0	..t t t c a g	att gaa..
24	116	..cta gcc AG	g t a t g c..	0.114	2	..c t g c a g	G gtc aca..
25	209	..atg gag gga G	g t g a g t..	0.077	1	..a g g g a g	GC aca gac..
26	133	..aag cac AA	g t g a g t..	0.435	2	..c t g c a g	G ttc tgc..
27	174	..act ttc AA	g t g a g t..	1.7	2	..c t g c a g	C atc aaa..
28	219	..ttc tac AG	g t g g g c..	1.7	2	..t c a c a g	G cct gag..
29	193	..ctc GAG	g t g g a g..	0.275	0	..c t g c a g	gcc tac..
30	85	..ggg aaa G	g t g t g t..	0.267	1	..g g g a a g	GG aag ggg..
31	129	..caa gca G	g t a a g c..	0.160	1	..t g c c a g	GC atc tcg..
32	283	..ttc ctc AG	g t a a g c..	0.174	2	..c t a c a g	C tac tgt..
33	142	..ctc CAG	g t g g g c..	1.3	0	..g t g c a g	gct gga..
34	167	..ata acg AG	g t c c g g..	0.334	2	..c t g c a g	G ctg agc..
35	178	..tgc AAG	g t a g g t..	0.350	0	..t t c t a g	gac atg..
36	196	..gca gaa G	g t g g g a..	0.337	1	..g a a c a g	GC aag gcc..
37	281	..cgg CAG	g t c a g t..	0.603	0	..c c g c a g	gtg ggt..
38	91	..gca tca G	g t a c g t..	0.309	1	..c c t c a g	CT gca gtt..
39	352	until poly(A) addition site					
		a g **g t** r a g t			y y n c **a g**		
mMTase consensus		79 84 100 97 95 63 84 45			79 79 74 100 100		
Rodent consensus		56 79 100 100 92 73 82 54			85 85 73 100 100		

Table 3.1. *The start codon corresponds to the third ATG (ATG3) in exon 1 (Gaudet et al., 1998). The stop codon (TAG) is located 39 bp into exon 39 and is followed bz a 311 bp untranslated region. The lines separating exon 30 and 31 show the separation between catalytic and regulatory domains. Intron sizes were determined by sequencing (GenBank accession numbers: AF 175410-175431, AF 234317-234318) and/or gel electrophoresis (Fig. 3.1). Intron phases are defined as the position of the intron within the codon.*

3.2 ALTERNATIVE *DNMT1* ISOFORMS DURING

GAMETOGENESIS

3.2.1 CLONING OF ALTERNATIVE DNMT1 ISOFORMS FROM TESTIS

In addition to the new 5' end exon reported by our group (Tucker et al., 1996) a small cDNA sequence of 155 bp detected in testis was found through searches in the EMBL database (EMBL accession number X 77486). Also, the presence of a ~ 1 kb larger *Dnmt1* transcript in testis had been reported (Trasler et al., 1992) part of which could correspond to this new EMBL sequence entry. Taken together these results indicated the presence of alternative *Dnmt1* isoforms in this tissue which interestingly is where profound changes in the overall DNA methylation level and pattern occur. Since the 155 bp cDNA sequence deposited in the database cannot account for the 1kb longer transcript found in testis, a screen for the missing cDNA sequence of this new *Dnmt1* isoform was started. Comparison of the EMBL sequence with the *Dnmt1* genomic sequence (GenBank accession number AF 175410; section 3.1; (Margot et al., 2000)) showed that the first 70 bp (depicted in dark blue in Fig.3.2) of the EMBL sequence are located between exons 1 and 2. This 70 bp part of the sequence is testis specific (thereafter referred to as testis-specific exon or t-exon) and it is located between nucleotides 1937-2005 (GenBank accession number AF 175411; (Margot et al., 2000)), the rest of the sequence up to 155 bp corresponds to exon 2 and 48 bp of exon 3 located between nucleotides 5242-5278 and 6041-6089 (GenBank accession number AF 175412; (Margot et al., 2000)) respectively. Taking a closer look at the *Dnmt1* gene sequence, between the end of exon 1 and the first nucleotide of the 70 bp testis specific sequence there is a distance of approximately 800 bp. Therefore, the testis specific sequence could either continuously extend at its 5' end and/or arise from alternative splicing. The information obtained from the genomic localization of the testis specific sequence allowed the design of oligos at different positions upstream of this 70 bp sequence and the screening of the 5' end of the *Dnmt1* gene by RT-PCR and RACE reactions (Fig.3.2).

Oligos located in the t-exon, exon 1 and exon 4 were selected for the RT-PCR and RACE reactions (Fig.3.2A) by using the Lasergene software program (Primer Selection, DNASTAR). Commercially available cDNA from mouse testis (Marathon cDNA, Clontech) was used for the RACE reactions. The RACE technique is a variation of the PCR method used to amplify cDNAs representing the region between a known point in a mRNA transcript and its 3' or 5' end (Frohman et al., 1988). A short internal stretch of sequence must already be known from the mRNA of interest and from this sequence gene-specific primers are chosen that are oriented in the direction of the missing sequence. Extension of the partial cDNAs from the unknown end of the message back to the known region is achieved using primers that anneal to the preexisting poly A tail (3' end) or to an -PCR, poly A RNA (100 ng) or total RNA (1 µg) were first converted into cDNA by reverse transcription using random primers. The subsequent PCR reactions were done with 100 ng of cDNA and the primer pairs indicated (Fig. 3.2B). The cycling conditions were according to the respective primer pair annealing temperatures. Fresh PCR products were then cloned into the pCRII vector (Invitrogen) and sequenced.

The size, sequence and location of the oligos used are shown in the table below:

Oligo Nr.	**Location**	**Used for:**	**Sequence**
243 (F)	exon1	RT-PCR	5' AAG ATG CCA GCG CGA ACA 3'
238 (R)	exon4	RT-PCR	5' TTG GCT TTT TGA GTG AGA 3'
242 (F)	exon1	RT-PCR	5' GGC TCG CTC CCG GAC CAT 3'
241 (R)	t-exon	RT-PCR	5' TGC CAT CCC TCA CTC CTC 3'
347 (R)	exon4	RACE	5' GGC TTT TTG AGT GAG AGT GTG TGT TCC G 3'
348 (R)	t-exon	RACE	5' TGC CAT CCC TCA CTC CTC GAA ATA AAC 3'

Two new isoforms were obtained (Fig. 3.2C): one was obtained by RT-PCR showing a new splicing variant and the second one was obtained by RACE analysis containing an extension of 316 bp at the 5' end of the t-exon. Further RACE analysis showed that the t-exon in this isoform is ~ 800 bp. The new splicing variant starts with exon 1 and continues with 88 bp of the t-exon (Fig. 3.2C). The exon-intron boundaries for this splicing variant match the GT/AG consensus (see also section 3.1). Due to the low abundance of this new isoform and the difficulty to obtain it, this isoform was not further investigated. The next sections therefore refer to the isoform obtained by RACE analysis.

While this work was in progress a testis-specific *Dnmt1* isoform was reported which corresponds to the one obtained by RACE, whose 5' boundary is approximately 80 nucleotides downstream of the 3' end of exon 1 of the ubiquitous mRNA (5.4 kb, Fig. 3.3). Furthermore in the ovaries a first exon specific to oocyte Mertineit, 1998 # 24 substitutes exon 1 and is located in the *Dnmt1* gene 7 kb upstream of exon 1 (Fig. 3.10).

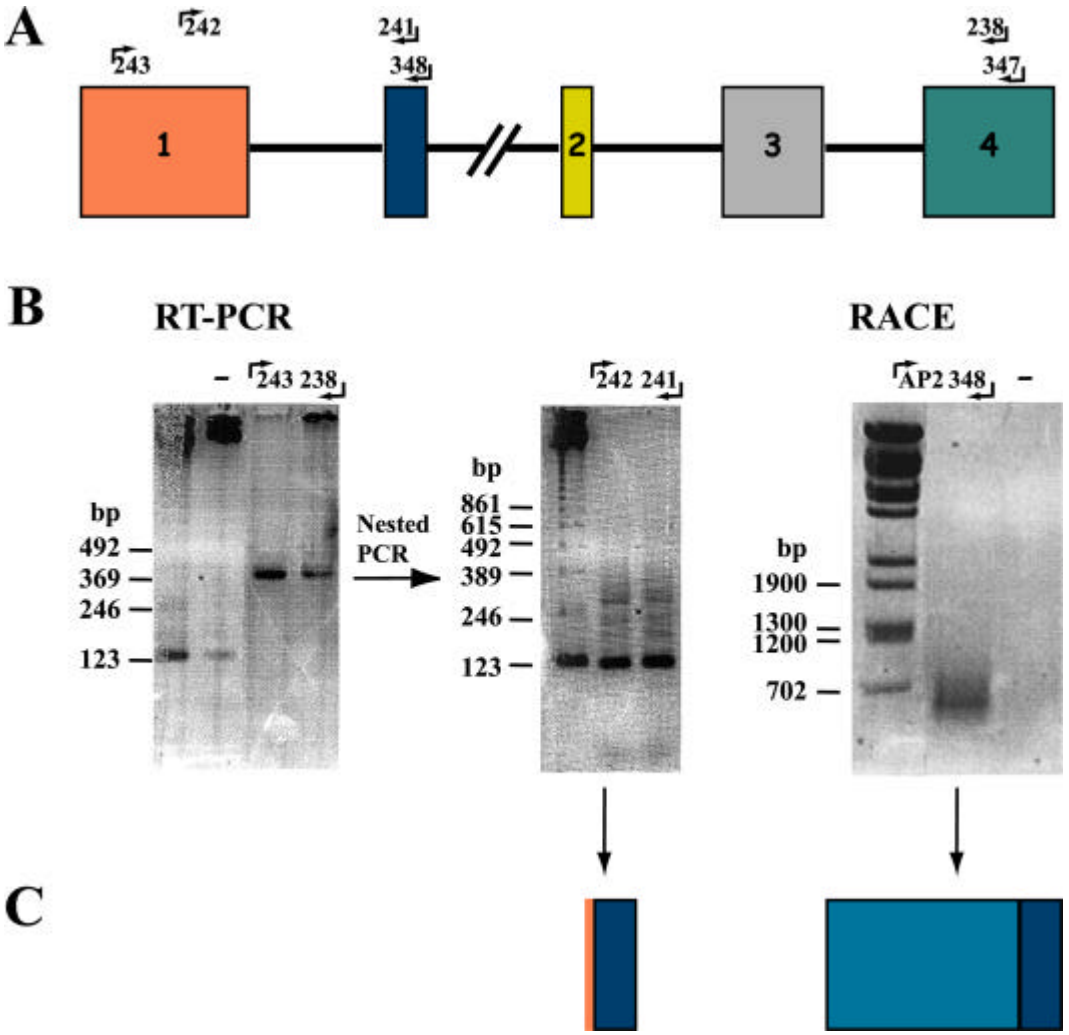


Fig. 3.2. Alternative dnmt1 isoforms in testis. (***A***) *Short Dnmt1 genomic scheme with the location of the oligos used.* (***B***) *Agarose gels showing the products from the RT-PCR (this reaction was carried out in duplicate) and the RACE reactions, the oligos used are indicated on top of each of the gels. The PCR control without cDNA is indicated with a minus sign (-).* (***C***) *Two new alternative Dnmt1 isoform are depicted. By RT-PCR, with random oligos and subsequent nested PCR with the oligos indicated, a new alternative splicing form was found starting with exon 1 and continuing with 88 nucleotides of the testis specific exon (dark blue). By RACE analysis an extension (light blue) of the 5' end of the t-exon was obtained using the oligos depicted in the panel B. A first RACE reaction was done with oligo 347 in combination with the adaptor primer 1 (AP1). An aliquot from this first reaction was then used in a second nested RACE reaction with oligo 348 in combination with adaptor primer 2 (AP2).*

3.2.2 EXPRESSION OF THE *DNMT1* ISOFORMS BY NORTHERN BLOT ANALYSIS

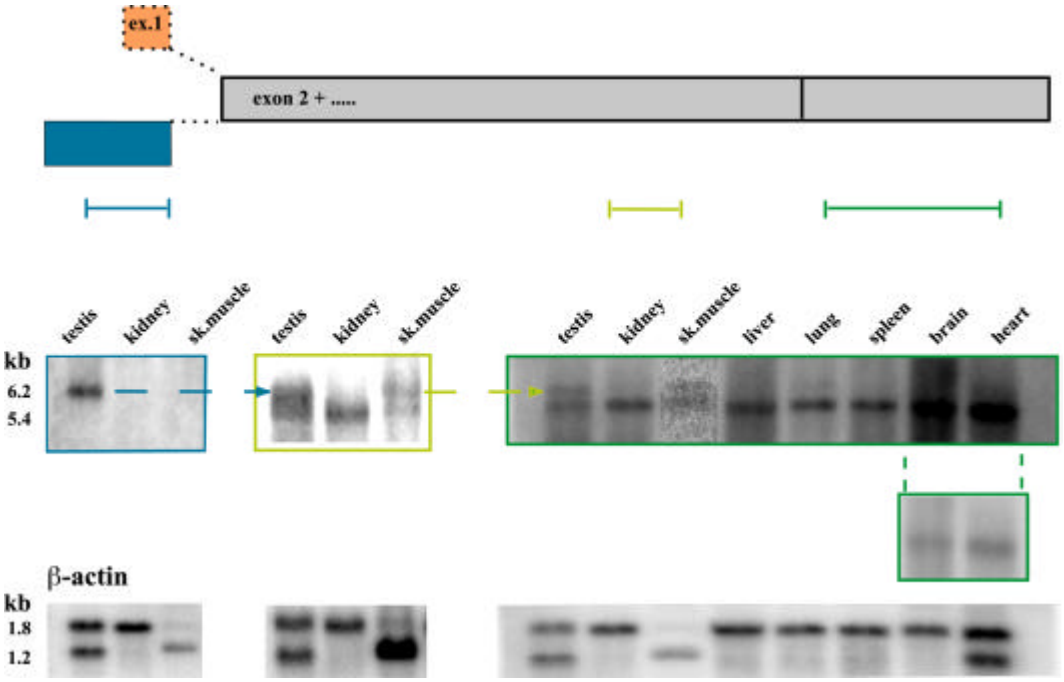
To find out the transcript size, tissue distribution and expression level of the *Dnmt1* isoforms Northern blot analysis was performed. Blots containing poly A RNA from mouse

testis (Clontech) were used. Three different cDNA probes were used for the Northern hybridizations are represented by color coding in Fig.3.3 and these are: a 377 bp (nucleotides 674 to 1050) and a 950 bp (nucleotides 2946 to 3895) corresponding to the N-terminal and Cterminal region of Dnmt1 respectively (GenBank accession number AF 162282; (Margot et al., 2000)) depicted in green in Fig.3.3. After hybridization with these probes, signals were detected and the blots were stripped by boiling with SDS solution (0.5% SDS) twice for 15 min, before being rehybridized with the next probe. Also a 396 bp probe (nucleotides 1621-2016, GenBank accession number AF 175411; (Margot et al., 2000)) was used, this fragment corresponds to the testis specific exon (Fig. 3.2C and also Fig. 3.3 depicted in blue). Finally, a mouse β–actin probe was hybridized to the blots to control for RNA loading. Labelling of the probes, hybridization, washes and exposure were done as described before (see Materials and Methods). The blots were exposed to a phosphorimager screen, scanned on a phosphorimager with the Image Quant software (Molecular Dynamics) and the results are shown in Fig. 3.3.

The use of two probes, depicted in green in Fig. 3.3, detected a *Dnmt1* transcript of 5.4 kb that is present in all tissues assayed. Its size was determined by comparing with the standard curve generated from the molecular weight markers. When differences in the amount of RNA loaded (as determined by probing the blot with β-actin) were taken into account, the 5.4 kb *Dnmt1* transcript was most abundant in heart and brain. Examination of the phosphorimages after longer exposure showed the presence of a slower migrating mRNA transcript of 6.2 kb that was present in testis and in skeletal muscle.

To investigate the structure of these longer transcripts blots were stripped and reprobed with a new probe containing testis specific sequences. If the slower mRNA transcript band (detected with the green depicted probes (Fig. 3.3)) contains the same sequences in testis and skeletal muscle, then it was expected to see this transcript in both organs when using a testis specific probe (depicted in blue in Fig. 3.3). However, the 6.2 kb transcript was detected by the testis specific probe only in testis and not in skeletal muscle. It might be that this slower migrating band in testis differs from the one present in skeletal muscle, or that the levels of the skeletal muscle transcript are so low that its presence is difficult to detect by Northern blot analysis. Higher amounts of RNA from skeletal muscle were tried but did not give interpretable results in part due to the lack of separation of high amounts of RNA on formaldehyde gels.

Fig. 3.3. Expression of Dnmt1 mRNA in different mouse tissues. *The figure shows Northern blots containing polyA RNA from the mouse tissues indicated. The blots were hybridized with three probes as indicated by the color coding. The black vertical bar indicates the separation between the N- and C-terminal domains. In the case of brain and heart tissue where Dnmt1 mRNA levels were high, a shorter exposure is shown underneath the corresponding blot. All blots were reprobed with a b-actin probe as a loading control.*



3.2.3 LOCALIZATION OF THE *DNMT1* ISOFORMS BY IN SITU HYBRIDIZATION ANALYSIS

The RACE and RT-PCR as well as the Northern blot analysis have shown that at least two major isoforms of *Dnmt1* are present, one that is ubiquitously expressed and the other that is detected in testis and in skeletal muscle. To analyze the localization of these isoforms at the cellular level, the in situ hybridization technique was used. The following is a brief introduction to the morphology and development of the testis to allow the identification of the mouse testis cells where the *Dnmt1* isoforms were localized.

The testis has two major components: the intertubular or interstitial compartment (interstitium) and the seminiferous tubule compartment. The interstitial compartment contains the blood and the lymphatic vessels. The seminiferous tubules are convoluted loops that have their two ends connected into the beginning of the rete testis. Although numerous convolutions are present in each loop, the tubules straighten between convolutions and travel largely in the long axis of the testis (Fig. 3.5C). As a result of this pattern of organization, a transverse histological section through the long axis of the testis can be used to visualize cross-sectioned tubules (Fig. 3.5C).

The adult mammalian testis contains male germ cells in all stages of development as well as several somatic cell types, including Sertoli, Leydig, and interstitial cells. The process of spermatogenesis can be divided into three main phases; a mitotic phase in which stem cells, called spermatogonia, replicate DNA and divide; a meiotic phase in which spermatocytes undergo genetic recombination and meiosis; and a differentiation phase, spermiogenesis, in which haploid cells, spermatids, acquire the specialized flagellum, Golgi, nucleus and mitochondria of the mature spermatozoon. During the differentiation or spermiogenesis, the spermatids transform themselves into œlls structurally equipped to reach and fertilize the egg (Fig. 3.4 and Fig. 3.5 C). The cells in the first meiotic prophase, primary spermatocytes, can be further subdivided according to the stages of

meiosis, preleptotene, leptotene, zygotene, pachytene and diplotene. The haploid phase can also be subdivided into a series of steps based on the morphology of the developing spermatids. The number and morphological characteristics of the steps differ slightly among mammals, in the case of mice 16 can be distinguished. In mice, cells in steps 1-8, round spermatids, are characterized by round, decondensed nuclei; cells in steps 9-early 12 are known as elongating spermatids, because the nuclear shape changes from round to elongated; and cells in the remaining steps are referred to in general as elongated spermatids (Fig. 3.4).

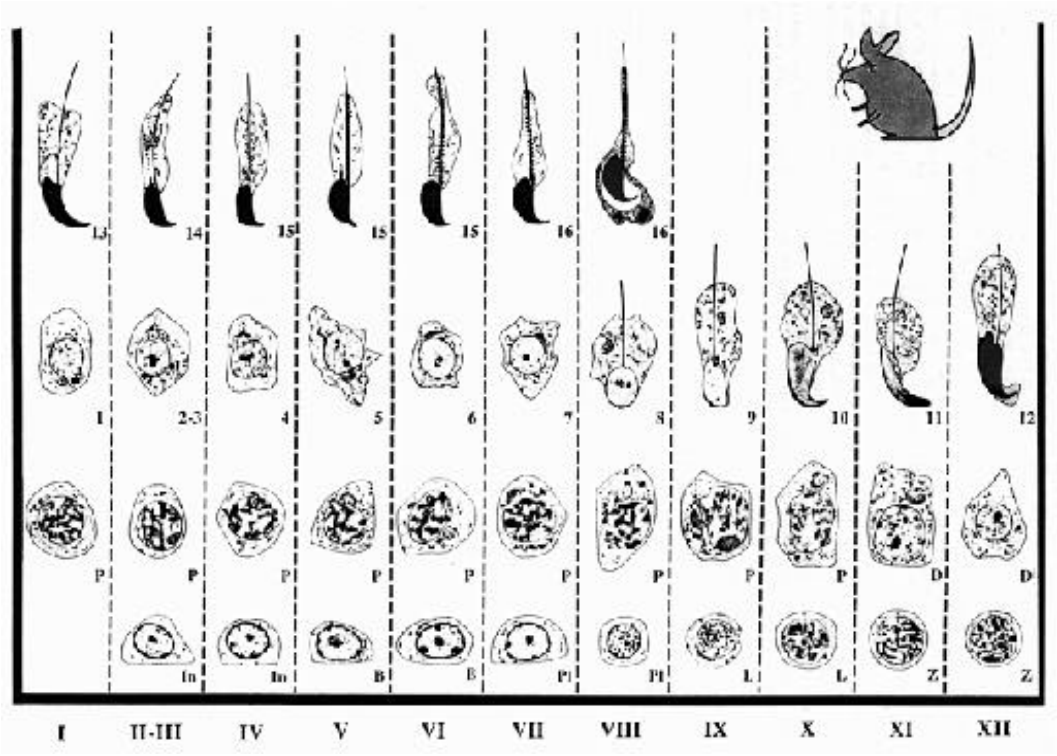
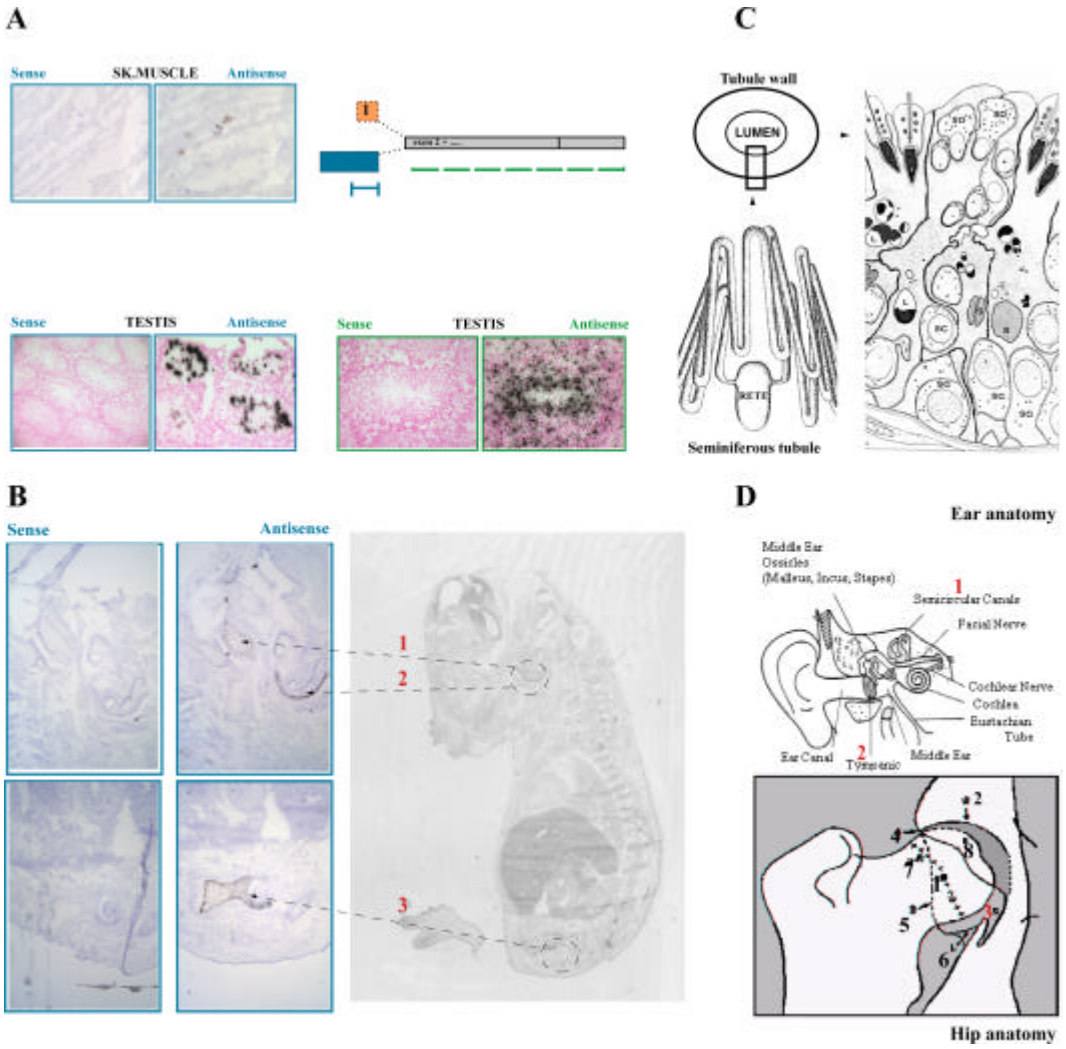


Fig.3.4. Map of the cycle of the seminiferous epithelium of the mouse (Russell et al., 1990). *The developmental progression of a cell is followed horizontally until the right hand border of the cycle map is reached. The cell progression continues at the left of the cycle map one row up. The vertical columns, designated by Roman numbers, depict cell associations (stages), meaning the different cell types that can be found in a testis cross section. In, intermediate spermatogonia; B, type B cells; Pl, preleptotene; L, leptotene; Z, zygotene; P, pachytene; D, diplotene.*

The Northern blot analysis indicated the presence of two (6.2 and 5.4 kb) *Dnmt1* transcripts in two tissues. The next step was to find out where these transcripts were localized at the cellular level in the organs that showed the new slower migrating transcript (6.2 kb) by Northern blot analysis. *Dnmt1* cDNA fragments (containing the areas depicted in Fig. 3.5A and B as blue and green) were cloned in the pCRII vector (Invitrogen). This vector contains T3 and T7 promoters in opposite sides of the multiple cloning site, allowing RNA to be transcribed *in vitro* from either strand of the insert. The cDNA fragments were cloned downstream from the T7 (depicted in blue) and from the T3 (depicted in green) RNA polymerase dependent promoter (Fig. 3.5A and B). The single stranded antisense (sense for the control) RNA probes were synthesized with the appropriate RNA polymerase and radioactive nucleotides by using an *in vitro* transcription

system (Promega). The radioactively labeled RNA transcripts were then hybridized with cryo sections of testis and skeletal muscle. The appropriate *Dnmt1* RNA probe was diluted in hybridization buffer and hybridized with the tissue sections overnight. The prehybridization times and the temperature for hybridization were optimized. The optimal conditions were found to be 37^{o}C for prehybridization for a minimal of 4 hours and 42^{o}C overnight for hybridization. Under these conditions the maximum sensitivity and the lowest background was obtained. Concurrently with the slides containing the antisense probe, adjacent tissue sections were also included containing a sense RNA probe as control for genomic DNA hybridization and unspecific binding. Signals were detected as described in Materials and Methods.

Fig. 3.5. Localization of Dnmt1 mRNA in different adult mouse tissues and embryo. *(**A**) and (**B**) show cryo sections of mouse testis, skeletal muscle and 13 day-old embryo. The hybridizations were done using probes as indicated by the color coding. Sagittal sections were used for skeletal muscle and embryo and for testis cross sections. Very weak signals (black dots) are seen in skeletal muscle (**A**) using the probe (depicted in blue) which also highlights the 6.2 kb band corresponding to the testis mRNA in the Northern blots (Fig. 3.3). When using the same probe in testis sections clear signals are localized on the cells closest to the lumen (spermatids). However, when using a probe that spans most of the cDNA (depicted in green), the signals are seen in most of the cells. The probe, depicted in blue, was used for embryo sections (**B**). Areas localized in the ear such as the lateral wall of right posterior semicircular canal (1) and right tubo tympanic recess (2) and also areas on the hip joint such as acetabular fossa (3) showed the presence of this transcript in embryo sections. The exposure times were 2 weeks for testis and embryo and 4 weeks for skeletal muscle. Negative controls (sense probe) were hybridized to adjacent sections. (**C**) Cross section of a seminiferous tubule. The schematic outline was taken from Clermont et al., (Clermont and Huckins, 1961) and the diagram showing the different types of cells on a cross section from testis was taken from Setchell et al., (Setchell et al., 1994). SG; spermatogonia, SC; primary spermatocyte, SD; spermatid, S; sertoli cell. (**D**)The ear and the hip anatomy are shown and the schematic outlines were taken from* http://www.bcm.tcm.edu/oto/studs/anat/tbone.html#ear *and* http://www.acay.com. au/~dissi/sbc/hipdys.htm, *respectively. Ear anatomy, (1) Semicircular canals and (2) Tympanic membrane are indicated in the figure. Hip anatomy, (**1**): Centre of the femoral head; (**2**): Cranial acetabular edge; (**3**): Acetabular fossa; (**4**): Cranial effective acetabular edge; (**5**): Dorsal acetabular edge; (**6**): Caudal acetabular edge; (**7**): Femoral head/neck reconturing; (**8**): Femoral ahead reconturing.*



When the probe which detects the ubiquitously expressed isoform (depicted in green in Fig. 3.5A) was hybridized to cross sections from mouse testis, all the cells that surrounded the lumen were labeled as shown by the black dots. When a probe that recognizes a sequence localized in the testis isoform was used (depicted in blue in Fig. 3.5A) the signals were more specific and localized to the spermatids which are the cells closer to the lumen. Adjacent cross sections that were labeled with a sense probe were used as a negative control and the absence of black dots in the whole section indicates the specificity of both antisense probes. The exposure times for the testis sections were ~10 days. A larger form of *Dnmt1* mRNA (which corresponds to the form depicted in blue in Fig. 3.5A) was shown to be present only in pachytene spermatocytes (Trasler et al., 1992) (see also Fig. 34). In these experiments isolated populations of male germ cells were used to prepare RNA for Northern blots . In these type of experiments it would be difficult to have pure populations of each cell type and thus contaminations by other cell types are difficult to avoid.

The same probe that hybridizes only to spermatids in testis sections (depicted in blue) was also used for skeletal muscle (sk. muscle) (Fig. 3.5A). Sagittal skeletal muscle sections were hybridized with the mentioned probe and after 4 weeks exposure very weak signals were detected, which suggests a low abundance of this transcript in this tissue. The sense control showed no signal.

To investigate whether this isoform is expressed during development and if so in which cell types, sagittal sections of whole mouse embryos (day 13) were hybridized to the probe corresponding to the t-exon (depicted in blue in Fig. 3.5A and B). The positive signals obtained were localized in cartilages of the ear and hip. Two areas of the ear where this isoform was localized were the lateral wall of right posterior semicircular canal and the right tubo tympanic recess (Fig. 3.5B; 1 and 2 respectively). Briefly, the ear can be divided into two areas the middle and the inner ear. It is in the inner ear where the cochlea which is the hearing part and the semicircular canals (Fig. 3.5D; ear anatomy) are located. The semicircular canals are not part of the hearing function, they are part of the balance system. For example, when a movement is done to turn the head, the fluid in the canal moves the tiny hairs of the nerve endings and bend them. Nerve impulses travel to the brain, giving information about the direction in which the head is moving. The other area in the embryo expressing the new isoform is the acetabular fossa which forms part of the hip (Fig. 3.5D, 3; hip anatomy). The new transcript has been detected in cartilages located in the ear (Fig. 3.5B, 1 and 2) and the area of the hip where this alternative transcript was present has also been reported to contain a cartilage covered part (Stener and Peterson, 1981). The skeletal muscle showed no detectable signals, in the 13 day-old embryo, when using the probe depicted in blue.

In summary, two *Dnmt1* transcripts are present: one that is ubiquitously expressed (5.4 kb) and a larger transcript (6.2 kb), which is present only in adult mouse testis and skeletal muscle. The slower migrating transcript has been further analyzed at the cellular level in testis sections and localized to the spermatids.

3.2.4 LOCALIZATION OF THE DNMT1 PROTEIN IN MALE AND FEMALE GERM CELLS

Overall DNA methylation changes dramatically during gametogenesis (see section 1.2.1) and therefore, in addition to the mRNA analysis, the level and localization of the Dnmt1 protein was studied in sections from ovaries and testis by immunohistochemistry.

In the female mouse the germ cells pass throught several oogonial divisions and reach the oocyte stage during embryonal development. About 3 days after birth the oocytes reach a static state known as the dictyate stage, which is maintained until a few hours before ovulation. A few days after birth some antra appear and the ovum reaches its maximum size at the time that antrum formation begins. Once oogenesis has begun, it continues in regular 4 1/2 to 5 day cycles throughout the reproductive life to about 12 or 14 months of age. A brief summary of the stages of oogenesis is depicted in the Fig. 3.6A. Briefly, after ovulation, the dominant follicle, reorganizes to become the corpus luteum. Thus, following rupture of the follicle, capillaries and fibroblasts from the surrounding stroma proliferate and penetrate the basal lamina and a rapid vascularization of the corpus luteus takes place. The granulosa cells present undergo morphological changes and this process is referred to as luteinization. These latter cells, the surrounding interstitial cells and the invading vasculature, give rise to a corpus luteum. It is this endocrine gland which is the major source of sex steroid hormones secreted by the ovary during the postovulatory phase of the cycle.

Ovaries and testis which were previously embedded in paraffin were sectioned and stained with the antibody pATH5, which recognizes multiple epitopes within the region between aminoacids 255-753 (Leonhardt et al., 1992; Li et al., 1992), which are present in both isoforms. The Dnmt1 protein is very abundant in oocytes as seen in Fig. 3.6B where mainly the cytoplasm is stained. In Fig. 3.6B a mature follicle is shown where the antrum (Fig. 3.6A and B) is already formed and the oocyte harbored in this cavity is intensively stained. The oocytes used were from adult female mice which explains the strong cytoplasmic staining. Mertineit et al. (Mertineit et al., 1998) recently showed

dramatic changes in the amount and localization of Dnmt1 during oogenesis. Growing oocytes showed intense nuclei and substantial cytoplasmic staining. At later stages of oocyte growth, Dnmt1 is no longer detectable in nuclei but accumulates to high levels in the cytoplasm. The Dnmt1 protein was also detected in the lutein cells (Fig. 3.6B).

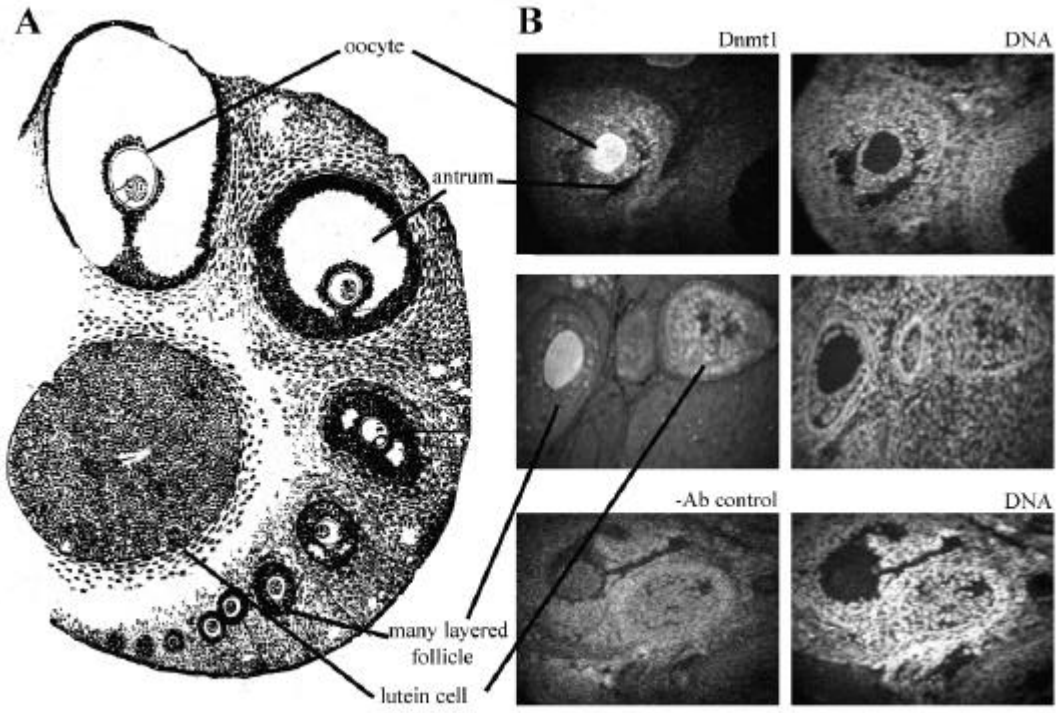


Fig. 3.6. Localization of Dnmt1 during oogenesis. *(**A**) Scheme of the development of an ovarian follicle is depicted (figure modified from R. Rugh). (**B**) Localization of the Dnmt1 protein by immunofluorescence staining with the rabbit polyclonal pATH5 antibody, detected with the Rhodamine conjugated anti-rabbit IgG antibody. The DNA counterstaining was performed with Hoechst 33258 to locate cell nuclei. The Dnmt1 protein is localized on the cytoplasm of the oocyte as indicated by the intense staining with the pATH5 antibody. – Ab control: represents the control without the first antibody.*

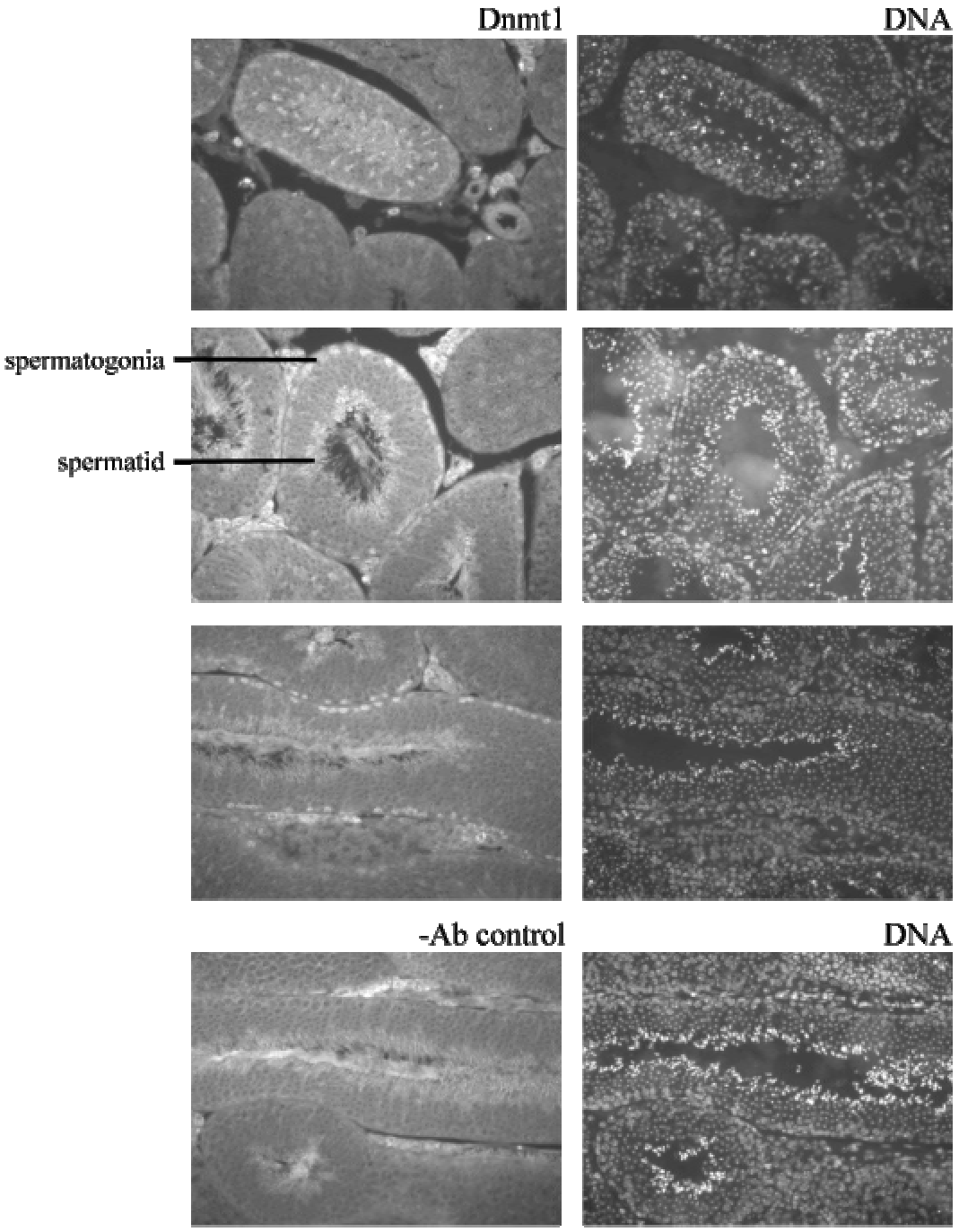


Fig. 3.7. Localization of Dnmt1 in seminiferous tubules. *The figure shows the localization of the Dnmt1 protein by inmunofluorescence staining with the rabbit polyclonal pATH5 antibody, detected with the Rhodamine conjugated anti rabbit IgG antibody. The DNA counterstaining was performed with Hoechst 33258 to locate the cell nuclei. The Dnmt1 protein is localized in spermatogonial cells which are the cells with a basal localization. The Dnmt1 is also seen in spermatids which are the cells closer to the lumen. –Ab control: indicates the control without the first antibody.*

The Dnmt1 staining in testis was localized to the spermatogonia, which are the mitotically active stem cells that lie at the basal side of the seminiferous epithelium, and in some primary spermatocytes, which are the next type of cells found in a testis cross section and are less basal than the spermatogonia cells. The spermatids also showed positive signals in some of the testis sections. These are the most specialized testis cells and correspond to a stage just previous to the formation of spermatozoon, in which the spermatids undergo morphological modification in a process called spermiogenesis (Fig. 3.4). However, the background level in the spermatids is high which makes it difficult to evaluate the signal observed in these cells. It is interesting that the new alternative *Dnmt1* transcript has also been localized to these cells by in situ hybridization (Fig. 3.5A), suggesting that this Dnmt1 protein staining (Fig. 3.7) could correspond to this isoform. That would contradict a recent report (Mertineit et al., 1998) proposing that this transcript is not translated.

3.3 EXPRESSION OF *DNMT1* ISOFORMS DURING MYOGENESIS

3.3.1 CLONING OF THE ALTERNATIVE *DNMT1* ISOFORM FROM SKELETAL MUSCLE

In mammals, levels of 5mC increase not only during gametogenesis but also after implantation when tissue specific patterns are established (Monk, 1990). During differentiation, when tissue specific genes begin to be expressed, the promoter regions of many of these genes are demethylated while the ones from inactive genes are methylated. One of the best characterized systems to study establishment of tissue specific gene expression is the skeletal muscle system, but conflicting data have been published concerning the role of DNA methylation during skeletal myogenesis. Thus on one hand, artificially induced demethylation seems to stimulate myogenic differentiation (Jones and Taylor, 1980), which is in accordance with the observation that the somatic *Dnmt1* isoform is downregulated during myogenesis (Liu et al., 1996). On the other hand ectopic overexpression of a truncated Dnmt1 protein in undifferentiated myoblasts leads to changes in gene expression and induction of myogenic differentiation (Takagi et al., 1995). The Northern blot analyses presented in Fig. 3.3 indicated the presence of two transcripts of 5.4 and 6.2 kb not only in testis but also in skeletal muscle. The shorter transcript corresponds to the ubiquitously expressed *Dnmt1* described before (Tucker et al., 1996; Yoder et al., 1996) but the longer transcript in muscle had not been previously reported and could correspond to an alternative isoform with potentially new properties. To isolate the cDNA corresponding to this alternative transcript, a combination of RACE and RT-PCR was used.

The *Dnmt1* cDNA has been shown to have 39 exons and to span over 56 kb (see section 3.1. Genomic structure). RT-PCR products from skeletal muscle RNA were compared with the previously described somatic isoform of *Dnmt1* (Tucker et al., 1996) from exon 5 to 37 to search for potential alternatively spliced exons (Fig. 3.8B). The oligos used in the PCR reactions as well as their exon localization and sequence are shown on the table below where the primer pairs are separated by horizontal lines.

Oligo Nr.	Location	Sequence
58 (F)	exon20	5' GAG CCA GGT AGA GAG TTA 3'
61 (R)	exon23	5' CAC CGC CAA GTT AGG ACA 3'
52 (F)	exon5	5' ACC ATC ACG GCT CAC TTC 3'
55 (R)	exon12	5' TCC TTG GGT CTC CGT TTG 3'
70 (F)	exon35	5' CCC TTC CGA ACC ATC ACC 3'
72 (R)	exon37	5' GCA GGC ACC AGG GGA TGA 3'
60 (F)	exon23	5' GGC GGT GAA GGA GGC AGA 3'
63 (R)	exon25	5' AGG GGC TTT GTA GAT GAC 3'
54 (F)	exon12	5' AAC GGA GAC CCA AGG AAG 3'
57 (R)	exon17	5' ACA CCC AGA AAA GTA GAG 3'
62 (F)	exon25	5' AAG GTC AAG GTC ATC TAC 3'
65 (R)	exon28	5' TTC TCC CCT TGA TGT AGT 3'
56 (F)	exon17	5' GGG CAC CTG TGT CCT GTC 3'
59 (R)	exon20	5' TGG GGG TCT CAT CAT CGT 3'
64 (F)	exon28	5' TCG GTC GGA TAA AAG AGA 3'
72 (R)	exon37	5' GCA GGC ACC AGG GGA TGA 3'

The oligos were 18 mers with low melting temperatures, thus annealing temperatures used in the PCR reactions were in the range of 40-45^{o}C.

The PCR product from the plasmid containing the ubiquitously expressed cDNA was always run on the left hand side from the PCR product obtained with cDNA from skeletal muscle as seen in part B from Fig. 3.8. Negative controls were also included in the reactions: one negative control for the PCR reaction which has all the PCR reagents except the cDNA and a RT negative control which includes all the reagents necessary to perform the RT reaction but not the enzyme (reverse transcriptase). The PCR products were in the range of 417-475 bp except for one that was 1500 bp long, due to the difficulties to find good primer pairs in the region. The negative controls showed no bands indicating the specificity of the PCR and RT reactions. No differences were found from exon 5 to exon 37 when screening for the presence of alternative exons by comparing RT-PCR products from skeletal muscle cDNA with a plasmid containing the ubiquitously expressed *Dnmt1* cDNA (Tucker et al., 1996).

The next step was to test whether new 3' or 5' exons are present in this alternative transcript by performing RACE reactions using gene-specific primers located in exon 37 and 4, respectively. The products obtained in both reactions were cloned and sequenced. From the 3' RACE reaction the sequences obtained (Fig. 3.8A) contained exons 37 to 39 from the previously described somatic isoform of *Dnmt1* and utilization of the first polyadenylation site (red box in Fig. 3.8A, see section 3.1. Genomic structure) (Margot et al., 2000), and therefore can not account for the longer mRNA found by Northern blot analysis in skeletal muscle (6.2 kb in Fig. 3.3). However, from the 5' RACE reactions two different size products were obtained and the respective sequences contained either exon 1 to 4 of the previously described ubiquitously expressed isoform (Tucker et al.,

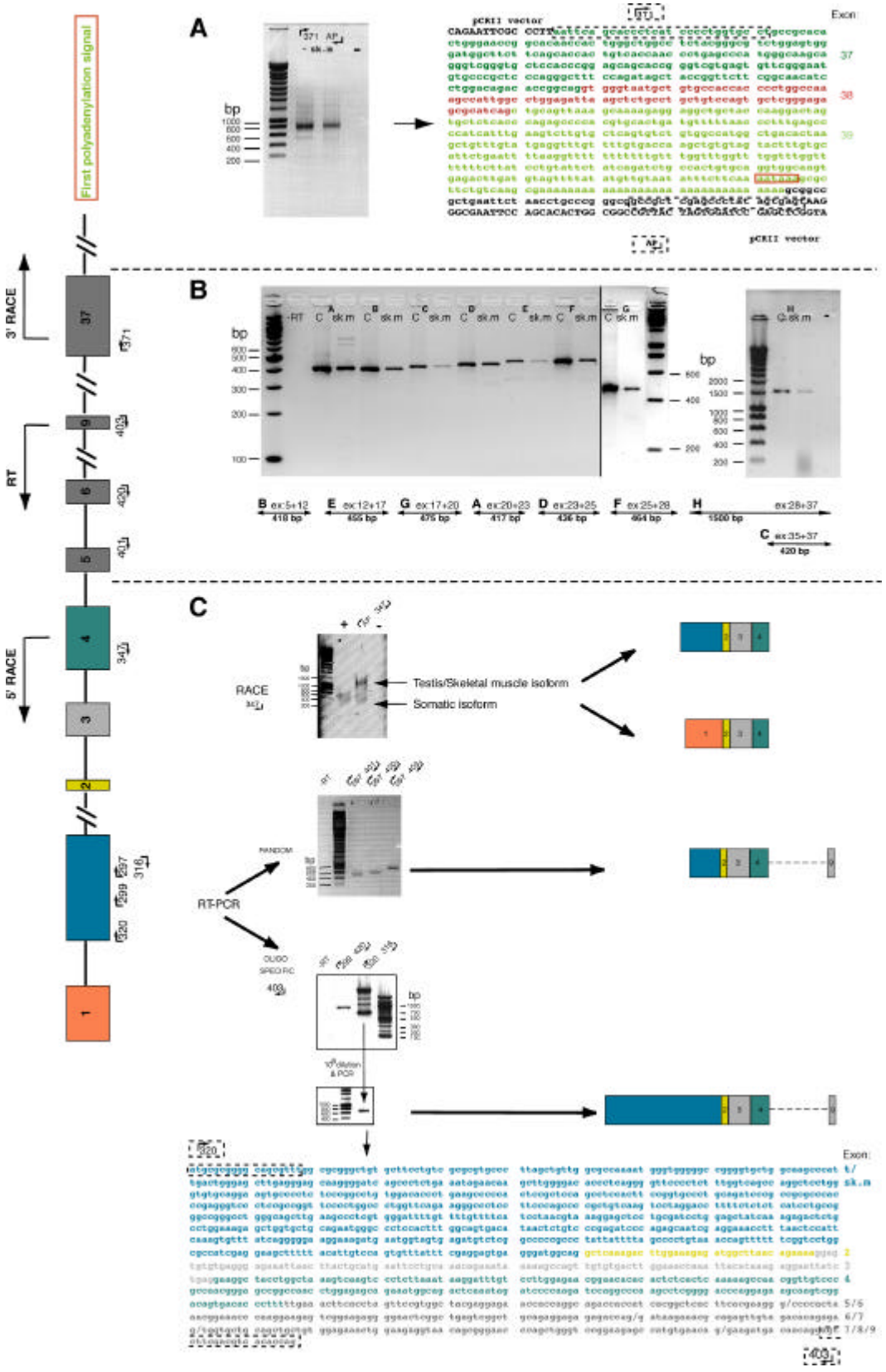
1996) or the t-exon followed by exons 2 to 4 (Fig. 3.8C). Since no longer RACE products were obtained, RT-PCR reactions using primers located at different positions in the genomic region upstream of the 5' end of the RACE product corresponding to the testis specific isoform were used to map the beginning of this exon in the skeletal muscle isoform. The oligos used in these RACE and PCR reactions as well as their exon localization and sequence are shown on the table below where primer pairs are separated by horizontal lines. Primers 371 and 347 were used in RACE reactions.

Oligo Nr.	**Location**	**Sequence**
371 (F)	exon1	5' AAT TCA GCA CCC TCA TCC CCT GGT GCC T 3'
347 (R)	exon4	5' GGC TTT TTG AGT GAG AGT GTG TGT TCC G
297 (F)	t-/sk.m ex.	5' GCC CCC GCC CTA TTA TTT TA 3'
401 (R)	exon5	5' TGA TGG TGG TCT GCC TGG T 3'
297 (F)	t-/sk.m ex.	5' GCC CCC GCC CTA TTA TTT TA 3'
420 (R)	exon4	5' TTC CCC TCT TCC GAC TCT TC 3'
297 (F)	t-/sk.m ex.	5' GCC CCC GCC CTA TTA TTT TA 3'
403 (R)	t-/sk.m ex.	5' CTG GTG TGA CGT CGA AGA CT 3'
320 (F)	t-/sk.m ex.	5' ATG CGC GGG GCA GCG TTT 3'
316 (R)	t-/sk.m ex.	5' TAA AAT AAT AGG GCG GGG GC 3'

The primers used for these reactions were 18-20 mers and the annealing temperature used in the PCR reactions was 55^{o}C. Some of the primer combinations and reaction products are shown in Fig. 3.8C .

The sequence of the longest 5' extension product amplified from skeletal muscle RNA is shown at the bottom of Fig. 3.8C (GenBank accession number AF 175432). This sequence turns out to be identical to the one identified in testis (section 3.2.1).

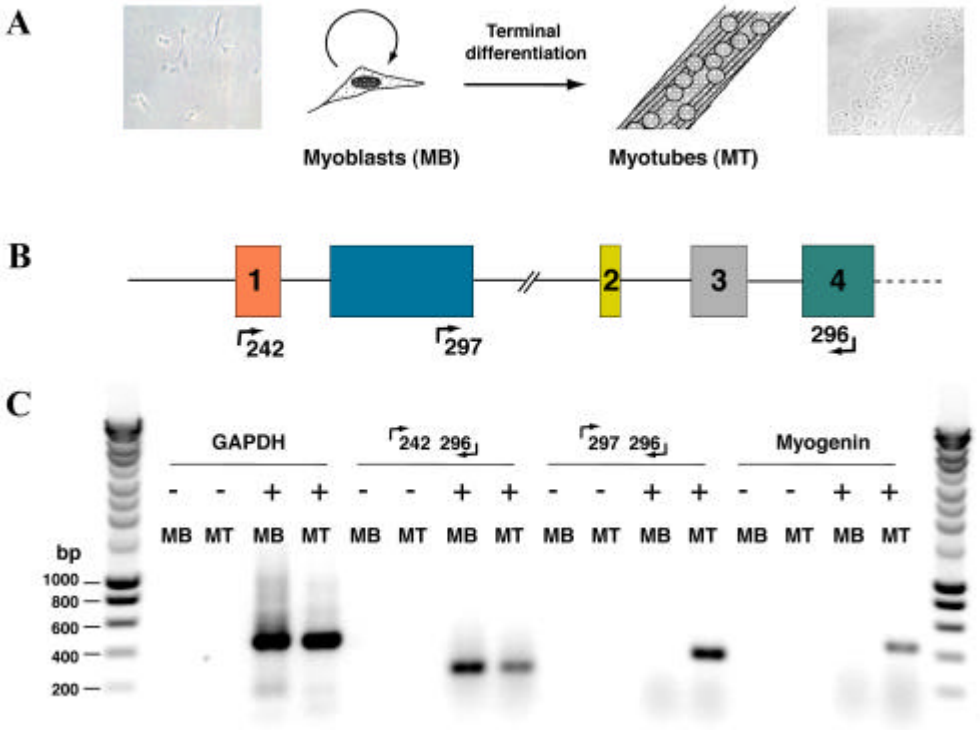
Fig. 3.8. Cloning of the skeletal muscle Dnmt1 isoform. *(**A**) 3' RACE was used to screen for differences at the 3' end which could account for the novel Dnmt1 transcript in skeletal muscle. A forward oligo in exon 37 (371) in combination with the adaptor primer (AP) were used. The PCR reaction was carried out in duplicate and the agarose shows the PCR products obtained whose size corresponds to approximately 800 bp. The PCR products were then cloned into the pCRII vector (sequence in capital letters) and sequenced. sk.m, corresponds to PCR product from skeletal muscle cDNA; -, represents the minus cDNA control PCR. The sequence obtained shows no new exons (from exon 37 to 39) and utilization of the first polyadenylation signal (indicated by a red box)(section 3.1 and (Margot et al., 2000)). (**B**) Screening for the presence of alternative exons by comparing RT-PCR products from skeletal muscle cDNA with a plasmid containing the ubiquitously expressed Dnmt1 cDNA. Skeletal muscle RNA was reverse transcribed using random primers. Different PCR reactions were done using oligo pairs as indicated, spanning the known Dnmt1 cDNA from exon 5 (ex. 5) to exon 37 (ex. 37). –RT, corresponds to the negative control without reverse transcriptase; -, corresponds to the negative control for the PCR reaction without cDNA template; C, PCR product from the plasmid; sk. m, PCR product from skeletal muscle cDNA. (**C**) To screen for new 5' end and/or alternative exons, a combination of oligos were used in 5' RACE and RT-PCR reactions using skeletal muscle mRNA. The corresponding location of the primers used and the 5' genomic structure (Margot et al., 2000) (see also Fig. 3.10) of the ubiquitous Dnmt1 (exon 1 to 9) as well as the testis specific isoform (indicated in blue) are depicted in the diagram. Reverse transcriptase reactions were done both with random and with and oligo (403) located in exon 9, and the RACE reaction with an oligo positioned in exon 4. –RT, as in (**B**); -, as in (**B**); +, positive PCR control. All the PCR products were cloned into pCRII and sequenced. From the RACE reactions two isoforms were obtained multiple times, which contain either exon 1 or the isoform isolated from testis (section 3.2.1). RT-PCR analysis using primers at different positions of the testis specific exon further confirmed the presence of this isoform in skeletal muscle mRNA. The sequence of the longest PCR product (GenBank accession number AF 175432) is shown underneath and the different exons are color coded as in the genomic structure.*


First polyadenylation signal
3' RACE
RT
5' RACE
A
B
C
bp
pCR11 vector
Exon
RACE
Testis/Skeletal muscle isoform
Somatic isoform
RT-PCR
RANDOM
OLIGO
SPECIFIC
Deletion & PCR

3.3.2 THE ALTERNATIVE MUSCLE ISOFORM IS EXPRESSED SPECIFICALLY IN DIFFERENTIATED MYOTUBES

The identification of this alternative *Dnmt1* isoform in skeletal muscle raises the question of when it is expressed during myogenic differentiation. To answer this question a well characterized *in vitro* differentiation system that is based on the ability of C2C12 mouse myoblasts (MB) to spontaneously differentiate into myotubes (MT) upon mitogen withdrawal (Fig. 3.9A) was used in this study. RNA from proliferating myoblasts as well as from differentiated myotubes was isolated and the expression of the two *Dnmt1* isoforms was analyzed by RT-PCR using upstream primers located either in exon 1 or in the t-/skeletal muscle specific exon (Fig. 3.9B). The ubiquitously expressed isoform containing exon 1 was present in both myoblasts and myotubes albeit at a lower level in differentiated cells (Fig. 3.9C). This is in agreement with a previous report showing a similar transcription rate but a shorter half-life of the mRNA for *Dnmt1* in differentiated mouse myotubes using a different myoblast cell line (Liu et al., 1996). The new isoform, however, was not detectable in undifferentiated myoblasts and was only expressed in differentiated myotubes (Fig. 3.9C). Equal amounts of input cDNA were present in both cases as shown by the similar amounts of GAPDH specific PCR fragment obtained using the same cDNA samples (Fig. 3.9 C). As a differentiation control primers specific for the mouse myogenin cDNA , a skeletal muscle specific transcription factor, which is expressed only after the onset of terminal differentiation (Andres and Walsh, 1996), were used. As expected, myogenin could only be amplified from myotube RNA (Fig. 3.9C). These results show that this alternative *Dnmt1* isoform is specifically expressed in myotubes during myogenesis.

Fig.3.9. *(**A**) The images and diagram above illustrate the myogenic differentiation system. Upon serum deprivation myoblasts (MB) fuse to form multinucleated myotubes (MT) which express several muscle-specific proteins such as the basic helix-loop-helix transcription factor myogenin. (**B**) Location of the oligos used for RT-PCR. (**C**) RT-PCR products (with and without reverse transcriptase, + and -) were analyzed in agarose gels. A 343 bp fragment was obtained in MT but not in MB using oligos located in the testis/skeletal muscle exon (depicted in blue) and in exon 4. For the ubiquitously expressed isoform, oligos located in exon 1 and exon 4 were used in the RT-PCR reaction, and the expected fragment of 264 bp was obtained in MB and at lower level in MT. RT-PCR with GAPDH specific oligos was used as a loading control giving rise to the 463 bp product seen at equal amounts in MB and MT. As a control for myogenic differentiation oligos that amplify a 450 bp fragment of the myogenin transcription factor were used and, as expected, a product is obtained in MT and not in MB.*



3.4 GENOMIC ORGANIZATION OF THE DNMT1 ISOFORMS

The most prominent *Dnmt1* isoforms seem to be generated by alternative transcriptional start sites. Therefore the 5' end of the *Dnmt1* gene was sequenced (GenBank acession number AF 175410-175413; about 26 kbp from the oocyte-specific exon to exon 9). In Fig. 3.10 the structure of the 5' end of the three *Dnmt1* isoforms identified so far is shown together with their genomic localization and their tissue distribution. Since these three alternative exons are not expressed in single tissues and to avoid confusion by exons identified in the future, they were renamed according to the chronological order of their identification: 1a, exon 1 (Tucker et al., 1996); 1b, spermatocytes (Mertineit et al., 1998), spermatids and skeletal muscle exon (this work); 1c, oocyte exon (Mertineit et al., 1998).

To identify CpG islands in the 5' end of the *Dnmt1* gene the CpG versus GpC sites (Fig. 3.10) are shown. The only CpG island identified between exon 1c and exon 3 is located around exon 1a, the start of the ubiquitously expressed isoform, and stretches till the start of tissue-specific exon 1b. In comparison, no CpG island could be identified around exon 3, which had previously been described as transcriptional start site of the ubiquitously expressed isoform (Rouleau et al., 1992). Since transcriptional start sites of housekeeping genes are typically associated with CpG islands, these results further support a transcriptional start of the ubiquitously expressed form with exon 1a.

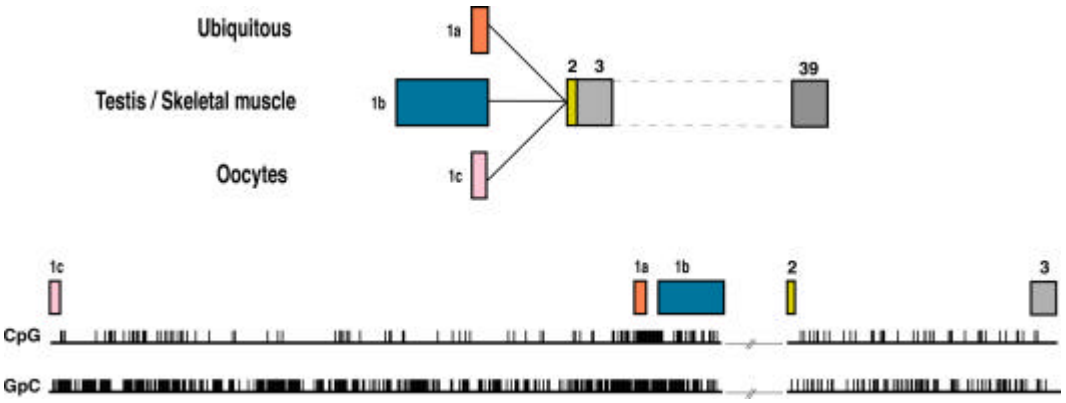


Fig. 3.10. The mouse Dnmt1 gene contains at least three alternative exons at the 5' end. *The diagram depicts the genomic structure of the 5' end of murine Dnmt1. Three alternative 5' exons have been identified: one is specific to the oocyte, one to the testis and skeletal muscle and one to somatic cells. The exons of Dnmt1 are represented to scale with respect to their size and relative position in the gene (GenBank accession number AF 175410-175412). The exons present in each isoform are indicated and are renamed following the chronological order of identification: 1a, exon 1 (Tucker et al., 1996); 1b, spermatocytes (Mertineit et al., 1998), spermatids and skeletal muscle exon (this work); 1c, oocyte exon (Mertineit et al., 1998). CpG and GpC incidence diagrams are plotted below to scale. The presence of a CpG island around exon 1a is a typical feature of housekeeping genes and fits well with the ubiquitous expression of this isoform.*

3.5 TRANSLATION EFFICIENCY OF THE TESTIS/ SKELETAL MUSCLE ISOFORM

Sequence analysis of the alternative *Dnmt1* transcript expressed in skeletal muscle and in testis revealed several short ORFs (Fig. 3.11). The first ATG is in-frame these short upstream ORFs has been previously proposed to prevent translation of this isoform in spermatocytes (Mertineit et al., 1998). The expression of an untranslatable mRNA, however, could hardly play an active role in the regulation of DNA methylation. Also, the 5'end of the *Dnmt1* transcript found in oocytes contains several short upstream ORFs (Mertineit et al., 1998) and yet is highly expressed *in vivo* (Fig. 3.6). Even the ubiquitously expressed form includes in the first exon (1a) three ATGs and only the third one is utilized *in vivo* (Gaudet et al., 1998).

To answer the question whether the skeletal muscle and testis *Dnmt1* transcript could be translated *in vivo*, a set of mammalian expression constructs containing the entire exon 1b as well as 5' truncated forms that contain only some of the short upstream ORFs were generated. The different *Dnmt1* pCR3.1 expression plasmids are designated 1a-9, 1b-9, 1b'-9, 1b''-9 and 3-9 (numbers correspond to exons present; see Fig. 3.10 for nomenclature of exons). To generate these constructs paired fragments of double stranded DNA, containing a common region that can be annealed after denaturation were used (see diagram on Fig.3.11A). Briefly, for each construct two paired PCR products of double stranded DNA were mixed with all the reagents necessary for a PCR reaction except the oligos. The oligo sequence and exon localization are indicated on the table below (oligo pairs are separated by horizontal lines). The oligo pairs were as follows: (584-582, 583-578) for 1a-9, (320-316, 579-578) for 1b-9, (579-581, 577-578) for 1b'-9. The mixture was then allowed to denature, anneal and extend for one cycle on a PCR

machine (Biometra). After this initial cycle, oligos flanking the newly formed double stranded DNA were added to the reaction and the PCR was carried out for another 30 cycles. The reverse oligo used in the latter reaction was the same in all the cases (616-R) which contains the FLAG sequence (indicated in bold letters). For the constructs 1b''-9 and 39, PCR reactions with oligos 577-578 and 325-616 respectively were carried out. The resulting PCR products were placed under the control of the CMV promoter by cloning into the pCR3.1 mammalian expression vector (Invitrogen).

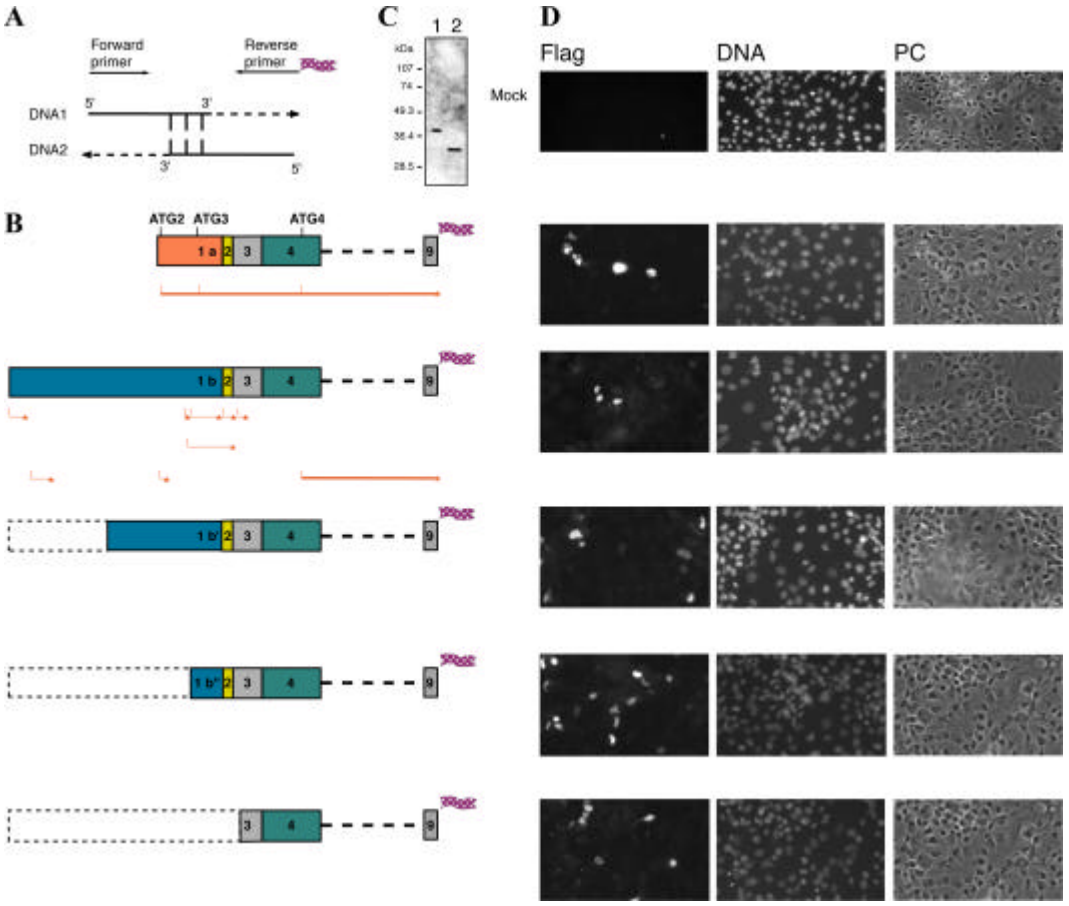
Oligo Nr.	Location	Sequence
584-F	exon1a	5´TAGCCAGGAGGTGTGGGTGCCTCCGTTGCGC 3´
582-R	exon 3	5´CTGTTTGCAGGAATTCATGCAGTAAGTTTAATTTCTCC 3´
583-F	exon 3	5´GCTCAAAGACTTGGAAAGAGATGGCTTAACAGAAAAG 3´
578-R	exon 9	5´CTT**CTTGTCATCGTCGTCCTTGTAGTC**TCTGGTGTGA 3´
320-F	exon1b	5' ATGCGCGGGGCAGCGTTT 3'
316-R	exon1b	5' TAAAATAATAGGGCGGGGGC 3'
579-F	exon1b	5' GCCCGGCTGTCAAGTCCTAGGACCTTTTCTCTCTCAT 3'
578-R	exon 9	5´CTT**CTTGTCATCGTCGTCCTTGTAGTC**TCTGGTGTGA 3´
579-F	exon1b	5' GCCCGGCTGTCAAGTCCTAGGACCTTTTCTCTCTCAT 3'
581-R	exon1b	5' TTCCCCTCTTCCGACTCTTCCTTGGGTTTCCGTTTAGT 3'
577-F	exon1b	5' GCCCCCGCCCTATTATTTTAGCCCCTGTAAACCAGT 3'
578-R	exon 9	5´CTT**CTTGTCATCGTCGTCCTTGTAGTC**TCTGGTGTGA 3´
325-F	exon 3	5'GAATCCCTGCAAACAGAAATAAAAAGCCAGTTGTGTGAC 3'
616-R	FLAG	5' CTT**CTTGTCATCGTCGTCCTTGTAGTC** 3'

All these constructs include exons 1b to 9 followed by the FLAG epitope sequence that is in frame with the Dnmt1 ORF and transcription is controlled by the CMV promoter (Fig. 3.11B). In these constructs the expression of the FLAG epitope is only possible when translation initiation takes place at ATG4 in exon 4. As a positive control, a construct containing most of exon 1a of the ubiquitously expressed form was generated. Plasmid DNA was prepared from these constructs and used for transfection of COS-7 cells. Three days after transfection, expression of the FLAG tagged Dnmt1 protein was analyzed by immunofluorescence staining and Western blotting using an anti-FLAG mouse monoclonal antibody. All the constructs were able to generate tagged Dnmt1 truncated protein in COS-7 cells as indicated by the positive FLAG signal in the immunofluorescence images shown in Fig. 3.11D. The non-specific background fluorescence under the same conditions is shown in the image of the stained mock transfected COS-7 cells.

Protein extracts from transfection experiments with a construct containing exon 1b'' and a control containing instead exon 1a were compared by Western blot analysis with antibodies against the FLAG epitope tag. Both constructs, 1a-9 and 1b''-9, gave rise to single bands on the western blot (Fig. 3.11C lanes 1 and 2 respectively). ATG4 is the only possible translation initiation site in the exon 1b constructs that is in frame with the FLAG epitope. These results clearly show that transcripts starting with exon 1b are

translatable and that a truncated Dnmt1 protein isoform is made, which could play an active role in the change of DNA methylation patterns during gametogenesis and myogenesis.

Fig. 3.11. The skeletal muscle Dnmt1 transcript is translatable. *Different expression constructs were designed to test whether the skeletal muscle isoform could generate a truncated Dnmt1 protein. (**A**)The strategy followed to generate some of the Dnmt1 expression constructs is depicted. In the first PCR cycle all the reagents necessary for a PCR reaction but the flanking oligos were included. The first cycle allows the annealing and extension of the overlapping area between the two denaturated DNAs (DNA1 and DNA2). In the next cycles the oligos were added and amplification of the desired area occurs. (**B**) The different epitope tagged Dnmt1 expression constructs used to transfect COS-7 cells; 1a-9, 1b-9, 3-9, 1b'-9 and 1b''-9 (b' and b''-9 correspond to nucleotides 651 to 1418, respectively from GenBank sequence accession number AF 175432). The exons included in each of the constructs are indicated (see Fig. 3.10 for exon nomenclature). The dotted lines in exon 1b, which is found in testis and skeletal muscle cDNA, represent the areas which were deleted in the specific construct. The small flags at the end of each construct represent the 8-amino acid FLAG coding sequence which was added to each of them. ORFs starting at ATG2 and ATG4 in exons 1a and 4 respectively of the ubiquitous Dnmt1 isoform cDNA (Gaudet et al., 1998) and finishing in the FLAG tag are shown in thick lines and the short ORFs are shown in thin lines. The short vertical lines correspond to ATGs in the particular reading frame and the stars represent stop codons. The different expression plasmids were used for transient transfection of COS-7 cells. (**C**) The two types of Dnmt1 protein bands obtained are shown in the Western blot and were detected with the anti-FLAG M2 mouse monoclonal antibody, using the Dnmt1 constructs 1a-9 (lane1) and 1b''-9 (lane2). (**D**) Expression of the FLAG tagged Dnmt1 protein (thick lines) was analyzed by immunofluorescence staining with the anti-Flag M2 mouse monoclonal antibody. The DNA counterstaining was performed with Hoechst 33258 and the corresponding phase contrast (PC) images are also shown. The mock transfection is depicted at the top using similar conditions.*


A
Forward primer
Reverse primer
DNA1
DNA2
5'
3'
C
1 2
kDa
B
ATG2 ATG3 ATG4
1 a 2 3 4 9
1 b 2 3 4 9
1 b' 2 3 4 9
3 4 9
D
Flag
DNA
PC
Mock

The results presented in this work indicate the presence of two transcripts of the *Dnmt1* gene whose expression varies substantially amongst the organs analyzed. Thus, high levels of expression of the ubiquitous transcript (5.4 kb) were detected in mouse brain and heart (Fig. 3.3). These results are in agreement with previous reports from Northern blot analyses in human tissues which showed that adult brain and heart express relatively large amounts of *Dnmt1* mRNA (Yen et al., 1992). Furthermore, a new slower migrating transcript (6.2 kb) was identified in testis and skeletal muscle (Fig. 3.3). A similar longer transcript had been previously reported in spermatogenic cells (pachytene spermatocytes) of the testis (Trasler et al., 1992). The size difference between the ubiquitous and the new transcript shown by Northern blots suggests the presence of an alternative transcriptional start site or an alternative splice isoform in testis and skeletal muscle. Indeed, by RT-PCR and RACE a new *Dnmt1* cDNA was isolated which differs in the first exon. Instead of the recently described exon 1 (Tucker et al., 1996; Yoder et al., 1996) an alternative 800 bp long exon located downstream in the gene is present in this isoform indicating a new transcriptional start site. While this work was in progress, a similar cDNA was independently isolated from testis (Mertineit et al., 1998).

The expression pattern of the new Dnmt1 isoform has been analyzed by in situ hybridization and found to be in the more specialized testis cells the haploid spermatids and at low level in skeletal muscle. The presence of a different Dnmt1 transcript in spermatids suggests that the function and regulation of Dnmt1 might be different in non-replicating (spermatids) and in replicating spermatogonial cells. Interestingly, the same isoform is also expressed in terminally differentiated post-mitotic myotubes (Fig. 3.9). During gametogenesis the overall DNA methylation level increases and is higher in the male than in the female gametes (Fig. 1.5) (Monk, 1990). The presence of a new Dnmt1 isoform in differentiating spermatids is suggestive of a role in these methylation pattern changes.

The expression of the new Dnmt1 isoform was also investigated by in situ hybridization in embryo sections and the new transcript was localized in cartilages of the ear and hip. Interestingly, different patterns of methylation have been described for specific collagen genes at different stages of differentiation and dedifferentiation of chondrocytes (Fernandez et al., 1985). However, it remains to be tested whether the alternative Dnmt1 isoform is involved in these methylation pattern changes.

Dramatic changes in genomic methylation occur during development in oogenesis and spermatogenesis (Fig. 1.5). Therefore, it was important to investigate Dnmt1 isoform expression and localization at the protein level in both organs, testis and ovaries. In testis sections, by immunofluorescence studies, spermatogonia as well as spermatids showed the presence of Dnmt1 protein (Fig. 3.7). However, in the case of spermatids the fluorescence background is high making the discrimination of signals less clear. Other studies (Mertineit et al., 1998) showed also high levels of Dnmt1 protein in spermatogonia, preleptotene and leptotene spermatocytes but in contrast to this work no protein was detected in spermatids and pachytene spermatocytes where the new transcript is made.

Ovaries from adult female mice were also analyzed for the presence of the Dnmt1 protein as shown in Fig. 3.6 where the Dnmt1 protein was localized mainly in the cytoplasm of mature oocytes and in the nucleus of lutein cells. Cytoplasmic retention of Dnmt1 protein in oocytes and preimplantation embryos has been previously reported (Carlson et al., 1992) and shown to be due to the presence of a retention signal in the N-terminal regulatory domain of the protein (Cardoso and Leonhardt, 1999). This active retention of Dnmt1 in the cytoplasm correlates well with the overall decrease in the genomic methylation level and thus is a likely mechanism to regulate DNA methylation by separating Dnmt1 from chromosomal DNA in the nucleus. Demethylation might occur by preventing maintenance methylation of newly synthesized DNA strands or by preventing remethylation of actively demethylated sites.

Changes in DNA methylation occur not only during gametogenesis but also during differentiation. However, conflicting data have been published concerning the exact role of DNA methylation during differentiation. Thus on one hand, artificially induced demethylation seems to stimulate myogenic differentiation (Jones and Taylor, 1980), which is in accordance with the observation that the somatic Dnmt1 isoform is downregulated during myogenesis (Liu et al., 1996). On the other hand, forced overexpression of a truncated Dnmt1 protein, was shown to cause de novo methylation in the MyoD gene, upregulate its expression and stimulate myogenesis (Takagi et al., 1995).

In this work, a new alternative transcript was identified not just in testis but also in skeletal muscle (Fig. 3.3) and shown to be specifically upregulated during myogenesis, while the ubiquitous transcript is downregulated (Fig. 3.9). These results could reconcile the above contradiction, in the sense that the major *Dnmt1* isoform responsible for maintenance of DNA methylation patterns is indeed downregulated but an alternative isoform with potentially different properties is upregulated.

The sequence analysis of this alternative *Dnmt1* isoform from skeletal muscle (Fig. 3.8) showed that it is actually identical to the one present in testis, which was published as untranslatable due to the presence of short upstream ORFs (Mertineit et al., 1998). These authors reported the inhibitory effect of upstream AUG codons based on the scanning model for translation initiation, in which the 43S pre-initiation complex enters at the 5'-cap structure of the mRNA and migrates in the 3'-direction to the first initiation codon that is present in an appropriate context (Kozak, 1991). Under some circumstances, however, an upstream AUG codon will reduce but not abolish initiation from downstream start codons. This happens when the first AUG codon is followed shortly by a terminator codon, creating a small open reading frame (ORF) at the 5' end of the mRNA. One possible explanation is that, after an 80S ribosome translates the 5'-mini ORF, the 40S ribosomal subunit remains bound to the mRNA, resumes scanning, and may reinitiate at another downstream AUG codon downstream (Kozak, 1991). The translational status of the alternative *Dnmt1* isoform was also analyzed experimentally (Trasler et al., 1992). These authors found that only ~15% of this *Dnmt1* transcript was associated with polysomes and that a large portion of this transcript exist as free mRNPs, suggesting that this alternative transcript is translationally regulated. However, if inhibition of translation takes place in this alternative isoform it should also take place in the somatic and in particular in the oocyte isoform which contains several upstream AUG codons. Instead the Dnmt1 protein is present at very high levels in the oocyte (Fig. 3.6). Furthermore, the suppressive influence of upstream initiation codons might be regulated. The cells have other mechanisms to counteract suppression by an upstream ORF. Thus, an upstream ORF can be bypassed either through `leaky scanning´ or through the use of an internal ribosome entry site (IRES) (Geballe and Morris, 1994).

When analysing these results together with the ones already published by other groups, it is clear that each experimental approach supports a different conclusion implicating technical factors. Thus, sedimentation profiles provide an accurate measure of translational activity but the most serious limitation is that the purified spermatogenic cells used for sucrose gradients contain a broad range of stages that could obscure translational differences in individual stages. This limitation could be overcome with the transillumination-assisted microdissection technique which permits the isolation of stages of spermatogenesis from freshly dissected seminiferous tubules (Mali et al., 1989). On the other hand, determining the stage of initial appearance of proteins and their mRNAs is subject to uncertainties caused by differences in the sensitivity of detection of mRNAs and proteins.

It is interesting to notice that two independent promoters are used in different tissues (oocytes versus testis and skeletal muscle) to generate two different transcripts, which however, lead to the expression of the same shorter Dnmt1 isoform (Fig. 3.10). The expression of an alternative Dnmt1 isoform during oogenesis, myogenesis and

spermatogenesis however does not prove an active role in this process, especially since both isoforms have very similar biochemical properties *in vitro* (Pradhan et al., 1997). The shorter isoform seems to be able to substitute for the longer isoform during differentiation and to restore DNA methylation patterns in *Dnmt1* null ES cells (Gaudet et al., 1998). Thus a specific role of the shorter isoform would require tissue-specific, interacting factors that could differentially interact with the Dnmt1 isoform and thus generate specificity. In this context is interesting to notice that the additional 118 amino acids of the longer isoform are highly conserved from sea urchin to human (Margot et al., 2000). Although the function of this amino acid stretch is not known, the presence of a putative leucine zipper suggests a role in protein-protein interactions which might be conserved throughout evolution.

First hints for a role of the short Dnmt1 isoform during development were obtained by serendipity. Since the complete somatic *Dnmt1* cDNA became known and available only in 1996 (Tucker et al., 1996; Yoder et al., 1996) a partial cDNA was used by other investigators for overexpression in myoblasts giving rise to a truncated Dnmt1 protein starting at ATG4, which is identical to the shorter isoform described here. The forced expression of this Dnmt1 isoform in myoblasts was shown to correlate with a high level of *de novo* methylation activity and to induce myogenic differentiation (Takagi et al., 1995). These experiments clearly show that the shorter Dnmt1 isoform from skeletal muscle cells can indeed induce myogenic differentiation.

In the present work the tissue specific expression of *Dnmt1* isoforms was shown which might play a role in the regulation of DNA methylation. However, further studies are needed involving selective *Dnmt1* gene knock outs and the examination of enzymatic properties of both Dnmt1 isoforms *in vivo*. The elucidation of the *Dnmt1* gene structure and the mapping of alternative transcriptional start sites carried out in this work make it now possible to study the transcriptional regulation and to characterize the tissue specific *Dnmt1*promoters.

Andres, V., and Walsh, K. (1996). Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis. J. Cell Biol. *132*, 657-666.

Antequera, F., Tamame, M., Villanueva, J. R., and Santos, T. (1984). DNA methylation in the fungi. J Biol Chem *259*, 8033-6.

Baylin, S. B., Makos, M., Wu, J. J., Yen, R. W., de Bustros, A., Vertino, P., and Nelkin, B. D. (1991). Abnormal patterns of DNA methylation in human neoplasia: potential consequences for tumor progression. Cancer Cells *3*, 383-90.

Bell, M. V., Hirst, M. C., Nakahori, Y., MacKinnon, R. N., Roche, A., Flint, T. J., Jacobs, P. A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U., and et al. (1991). Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell *64*, 861-6.

Bestor, T. H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. Embo J *11*, 2611-2617.

Bestor, T. H. (1988). Cloning of a mammalian DNA methyltransferase. Gene *74*, 9-12.

Bestor, T. H. (1990). DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. Philos Trans R Soc Lond B Biol Sci *326*, 179-87.

Bestor, T. H. (1998). Gene silencing. Methylation meets acetylation. Nature *393*, 311-2.

Bestor, T. H., Hellewell, S. B., and Ingram, V. M. (1984). Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. Mol Cell Biol *4*, 1800-6.

Bhattacharya, S. K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999). A mammalian protein with specific demethylase activity for mCpG DNA. Nature *397*, 579-83.

Bird, A. (1992). The essentials of DNA methylation. Cell *70*, 5-8.

Bird, A., Taggart, M., Frommer, M., Miller, O. J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res *7*, 1513-23.

Buschhausen, G., Wittig, B., Graessmann, M., and Graessmann, A. (1987). Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. Proc Natl Acad Sci U S A *84*, 1177-81.

Cardoso, M. C., and Leonhardt, H. (1999). Differentiation, development and programmed cell death. In The Molecular Basis of Cell Cycle and Growth Control, G. Stein, R. Baserga, D. Denhardt and A. Giordano, eds. (New York: John Wiley & Sons), pp. 305-347.

Cardoso, M. C., and Leonhardt, H. (1999). DNA methyltransferase is actively retained in the cytoplasm during early development. J Cell Biol *147*, 25-32.

Carlson, L. L., Page, A. W., and Bestor, T. H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev *6*, 2536-2541.

Carrera, A., Gerton, G. L., and Moss, S. B. (1994). The major fibrous sheath polypeptide of mouse sperm: structural and functional similarities to the A-kinase anchoring proteins. Dev Biol *165*, 272-84.

Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G. G., and Wells, R. D. (1987). Cloning, sequencing, in vivo promoter mapping, and expression in Escherichia coli of the gene for the HhaI methyltransferase. J Biol Chem *262*, 4770-7.

Chaillet, J. R., Vogt, T. F., Beier, D. R., and Leder, P. (1991). Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell *66*, 77-83.

Cheng, X. (1995). DNA modification by methyltransferases. Curr Opin Struct Biol *5*, 4-10.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., and Roberts, R. J. (1993). Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell *74*, 299-307.

Chernov, A. V., Vollmayr, P., Walter, J., and Trautner, T. A. (1997). Masc2, a C5-DNA-methyltransferase from Ascobolus immersus with similarity to methyltransferases of higher organisms. Biol Chem *378*, 1467-73.

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K., and McCann, P. P. (1996). S-Adenosylmethionine and methylation. Faseb J *10*, 471-80.

Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G., and Li, B. F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. Proc Natl Acad Sci U S A *81*, 1991-5.

Clark, S. J., Harrison, J., and Frommer, M. (1995). CpNpG methylation in mammalian cells. Nat Genet *10*, 20-7.

Clermont, Y., and Huckins, C. (1961). Microscopic anatomy of the sex cords and seminiferous tubules in growing andadult male albino rats. Americal Journal of Anatomy *108*, 79-97.

Cooper, D. N., and Youssoufian, H. (1988). The CpG dinucleotide and human genetic disease. Hum Genet *78*, 151-5.

Coulondre, C., Miller, J. H., Farabaugh, P. J., and Gilbert, W. (1978). Molecular basis of base substitution hotspots in Escherichia coli. Nature *274*, 775-80.

Counts, J. L., and Goodman, J. I. (1995). Alterations in DNA methylation may play a variety of roles in carcinogenesis. Cell *83*, 13-15.

Cross, S. H., and Bird, A. P. (1995). CpG islands and genes. Curr. Opin. Genet. Dev. *5*, 309-314.

Crowther, P. J., Doherty, J. P., Linsenmeyer, M. E., Williamson, M. R., and Woodcock, D. M. (1991). Revised genomic consensus for the hypermethylated CpG island region of the human L1 transposon and integration sites of full length L1 elements from recombinant clones made using methylation-tolerant host strains. Nucleic Acids Res *19*, 2395-401.

Dagert, M., and Ehrlich, S. D. (1979). Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene *6*, 23-8.

de Bustros, A., Nelkin, B. D., Silverman, A., Ehrlich, G., Poiesz, B., and Baylin, S. B. (1988). The short arm of chromosome 11 is a "hot spot" for hypermethylation in human neoplasia. Proc Natl Acad Sci U S A *85*, 5693-7.

Doerfler, W., Toth, M., Kochanek, S., Achten, S., Freisem-Rabien, U., Behn-Krappa, A., and Orend, G. (1990). Eukaryotic DNA methylation: facts and problems. FEBS Lett *268*, 329-33.

Feinberg, A. P. (1993). Genomic imprinting and gene activation in cancer. Nature Genet. *4*, 110-113.

Feinberg, A. P., Gehrke, C. W., Kuo, K. C., and Ehrlich, M. (1988). Reduced genomic 5-methylcytosine content in human colonic neoplasia. Cancer Res *48*, 1159-61.

Fernandez, M. P., Young, M. F., and Sobel, M. E. (1985). Methylation of type II and type I collagen genes in differentiated and dedifferentiated chondrocytes. J Biol Chem *260*, 2374-8.

Frohman, M. A., Dush, M. K., and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A *85*, 8998-9002.

Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K. C., Gehrke, C. W., and Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res *11*, 6883-94.

Gaudet, F., Talbot, D., Leonhardt, H., and Jaenisch, R. (1998). A short DNA methyltransferase isoform restores methylation In vivo. J Biol Chem *273*, 32725-9.

Geballe, A. P., and Morris, D. R. (1994). Initiation codons within 5'-leaders of mRNAs as regulators of translation. Trends Biochem Sci *19*, 159-64.

Gerlt, J. A., and Gassman, P. G. (1993). Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl-transfer reactions, and displacement reactions of phosphodiesters. Biochemistry *32*, 11943-52.

Giovannucci, E., Stampfer, M. J., Colditz, G. A., Rimm, E. B., Trichopoulos, D., Rosner, B. A., Speizer, F. E., and Willett, W. C. (1993). Folate, methionine, and alcohol intake and risk of colorectal adenoma. J Natl Cancer Inst *85*, 875-84.

Glickman, J. F., Pavlovich, J. G., and Reich, N. O. (1997). Peptide mapping of the murine DNA methyltransferase reveals a major phosphorylation site and the start of translation. J Biol Chem *272*, 17851-7.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell *23*, 175-82.

Gomperts, M., Garcia-Castro, M., Wylie, C., and Heasman, J. (1994). Interactions between primordial germ cells play a role in their migration in mouse embryos. Development *120*, 135-41.

Green, P. M., Montandon, A. J., Bentley, D. R., Ljung, R., Nilsson, I. M., and Giannelli, F. (1990). The incidence and distribution of CpG----TpG transitions in the coagulation factor IX gene. A fresh look at CpG mutational hotspots. Nucleic Acids Res *18*, 3227-31.

Greenblatt, M. S., Bennett, W. P., Hollstein, M., and Harris, C. C. (1994). Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res *54*, 4855-78.

Greger, V., Passarge, E., Hopping, W., Messmer, E., and Horsthemke, B. (1989). Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet *83*, 155-8.

Groudine, M., and Conkin, K. F. (1985). Chromatin structure and de novo methylation of sperm DNA: implications for activation of the paternal genome. Science *228*, 1061-8.

Gruenbaum, Y., Cedar, H., and Razin, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. Nature *295*, 620-2.

Gruenbaum, Y., Naveh-Many, T., Cedar, H., and Razin, A. (1981). Sequence specificity of methylation in higher plant DNA. Nature *292*, 860-2.

Hatada, I., Ohashi, H., Fukushima, Y., Kaneko, Y., Inoue, M., Komoto, Y., Okada, A., Ohishi, S., Nabetani, A., Morisaki, H., Nakayama, M., Niikawa, N., and Mukai, T. (1996). An imprinted gene p57KIP2 is mutated in Beckwith-Wiedemann syndrome. Nat Genet *14*, 171-3.

Herman, J. G., Latif, F., Weng, Y., Lerman, M. I., Zbar, B., Liu, S., Samid, D., Duan, D. S., Gnarra, J. R., Linehan, W. M., and et al. (1994). Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A *91*, 9700-4.

Holliday, R., and Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. Science *187*, 226-32.

Howlett, S. K., and Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. Development *113*, 119-127.

Hung, M. S., Karthikeyan, N., Huang, B., Koo, H. C., Kiger, J., and Shen, C. J. (1999). Drosophila proteins related to vertebrate DNA (5-cytosine) methyltransferases. Proc Natl Acad Sci U S A *96*, 11940-5.

Huynh, H., Alpert, L., and Pollak, M. (1996). Silencing of the mammary-derived growth inhibitor (MDGI) gene in breast neoplasms is associated with epigenetic changes. Cancer Res *56*, 4865-70.

Ingrosso, D., Fowler, A. V., Bleibaum, J., and Clarke, S. (1989). Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs for protein, DNA, RNA, and small molecule S-adenosylmethionine-dependent methyltransferases. J Biol Chem *264*, 20131-9.

Jones, P. A. (1986). DNA methylation and cancer. Cancer Res *46*, 461-6.

Jones, P. A. (1984). Gene activation by 5-azacytidine (New York: Springer-Verlag).

Jones, P. A., Rideout, W. M. d., Shen, J. C., Spruck, C. H., and Tsai, Y. C. (1992). Methylation, mutation and cancer. Bioessays *14*, 33-6.

Jones, P. A., and Taylor, S. M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. Cell *20*, 85-93.

Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet *19*, 187-91.

Jost, J. P., and Jost, Y. C. (1994). Transient DNA demethylation in differentiating mouse myoblasts correlates with higher activity of 5-methyldeoxycytidine excision repair. J Biol Chem *269*, 10040-10043.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev *6*, 705-714.

Kafri, T., Gao, X., and Razin, A. (1993). Mechanistic aspects of genome-wide demethylation in the preimplantation mouse embryo. Proc Natl Acad Sci U S A *90*, 10558-62.

Kass, S. U., Landsberger, N., and Wolffe, A. P. (1997). DNA methylation directs a time-dependent repression of transcription initiation. Curr Biol *7*, 157-65.

Kautiainen, T. L., and Jones, P. A. (1986). DNA methyltransferase levels in tumorigenic and nontumorigenic cells in culture. J Biol Chem *261*, 1594-8.

Keshet, I., Lieman-Hurwitz, J., and Cedar, H. (1986). DNA methylation affects the formation of active chromatin. Cell *44*, 535-43.

Kleene, K. C. (1989). Poly(A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse. Development *106*, 367-73.

Klimasauskas, S., Nelson, J. L., and Roberts, R. J. (1991). The sequence specificity domain of cytosine-C5 methylases. Nucleic Acids Res *19*, 6183-90.

Klimasauskas, S., Steponaviciene, D., Maneliene, Z., Petrusyte, M., Butkus, V., and Janulaitis, A. (1990). M.Smal is an N4-methylcytosine specific DNA-methylase. Nucleic Acids Res *18*, 6607-9.

Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V., and Janulaitis, A. (1989). Sequence motifs characteristic of DNA[cytosine-N4]methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. Nucleic Acids Res *17*, 9823-32.

Kochanek, S., Renz, D., and Doerfler, W. (1993). Differences in the accessibility of methylated and unmethylated DNA to DNase I. Nucleic Acids Res *21*, 5843-5845.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem *266*, 19867-70.

Kumar, S., Cheng, X., Pflugrath, J. W., and Roberts, R. J. (1992). Purification, crystallization, and preliminary X-ray diffraction analysis of an M.HhaI-AdoMet complex. Biochemistry *31*, 8648-53.

Laird, P. W., and Jaenisch, R. (1994). DNA methylation and cancer. Hum Mol Genet *3*, 1487-95.

Lapidus, R. G., Ferguson, A. T., Ottaviano, Y. L., Parl, F. F., Smith, H. S., Weitzman, S. A., Baylin, S. B., Issa, J. P. J., and Davidson, N. E. (1996). Methylation of estrogen and progesterone receptor gene 5' CpG islands correlates with lack of estrogen and progesterone receptor gene expression in breast tumors. Clin Cancer Res *2*, 805-10.

Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

Leonhardt, H., and Bestor, T. H. (1993). Structure, function and regulation of mammalian DNA methyltransferase. In DNA Methylation: Molecular Biology and Biological Significance, J. P. Jost and H. P. Saluz, eds. (Basel: Switzerland: Birkäuser Verlag), pp. 109-119.

Leonhardt, H., Page, A. W., Weier, H. U., and Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865-73.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905-14.

Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-26.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709-15.

Little, M., and Wainwright, B. (1995). Methylation and p16: suppressing the suppressor [comment]. Nat Med *1*, 633-4.

Liu, Y., Oakeley, E. J., Sun, L., and Jost, J. P. (1998). Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. Nucleic Acids Res *26*, 1038-1045.

Liu, Y., Sun, L., and Jost, J. P. (1996). In differentiating mouse myoblasts DNA methyltransferase is posttranscriptionally and posttranslationally regulated. Nucleic Acids Res *24*, 2718-2722.

Makos, M., Nelkin, B. D., Chazin, V. R., Cavenee, W. K., Brodeur, G. M., and Baylin, S. B. (1993). DNA hypermethylation is associated with 17p allelic loss in neural tumors. Cancer Res *53*, 2715-8.

Malagnac, F., Wendel, B., Goyon, C., Faugeron, G., Zickler, D., Rossignol, J. L., Noyer-Weidner, M., Vollmayr, P., Trautner, T. A., and Walter, J. (1997). A gene essential for de novo methylation and development in Ascobolus reveals a novel type of eukaryotic DNA methyltransferase structure. Cell *91*, 281-90.

Mali, P., Kaipia, A., Kangasniemi, M., Toppari, J., Sandberg, M., Hecht, N. B., and Parvinen, M. (1989). Stage-specific expression of nucleoprotein mRNAs during rat and mouse spermiogenesis. Reprod Fertil Dev *1*, 369-82.

Mandel, M., and Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. J Mol Biol *53*, 159-62.

Margot, J. B., Aguirre-Arteta, A. M., Di Giacco, B. V., Pradhan, S., Roberts, R. J., Cardoso, M. C., and Leonhardt, H. (2000). Structure and function of the mouse DNA methyltransferase gene. Dnmt1 shows a tripartite structure. Journal of Molecular Biology *297*, 293-300.

Marinus, M. G. (1984). DNA Methylation. Biochemical and Biological Significance, A. Rich, ed. (New York: Springer-Verlag).

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., and Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499-507.

Mertineit, C., Yoder, J. A., Taketo, T., Laird, D. W., Trasler, J. M., and Bestor, T. H. (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development *125*, 889-97.

Mi, S., and Roberts, R. J. (1992). How M.MspI and M.HpaII decide which base to methylate. Nucleic Acids Res *20*, 4811-6.

Monk, M. (1990). Changes in DNA methylation during mouse embryonic development in relation to Xchromosome activity and imprinting. Philos Trans R Soc Lond B Biol Sci *326*, 299-312.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development *99*, 371-382.

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-9.

Nicholls, R. D., Saitoh, S., and Horsthemke, B. (1998). Imprinting in Prader-Willi and Angelman syndromes. Trends Genet *14*, 194-200.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M. F., and Mandel, J. L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science *252*, 1097-102.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-57.

Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases [letter]. Nat Genet *19*, 219-20.

Okano, M., Xie, S., and Li, E. (1998). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res *26*, 2536-2540.

Paroush, Z., Keshet, I., Yisraeli, J., and Cedar, H. (1990). Dynamics of demethylation and activation of the alpha-actin gene in myoblasts. Cell *63*, 1229-1237.

Pfeifer, G. P., Steigerwald, S. D., Hansen, R. S., Gartler, S. M., and Riggs, A. D. (1990). Polymerase chain reaction-aided genomic sequencing of an X chromosome- linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. Proc Natl Acad Sci U S A *87*, 8252-6.

Pieretti, M., Zhang, F. P., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., and Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. Cell *66*, 817-22.

Pradhan, S., Talbot, D., Sha, M., Benner, J., Hornstra, L., Li, E., Jaenisch, R., and Roberts, R. J. (1997). Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase. Nucleic Acids Res *25*, 4666-73.

Proffitt, J. H., Davie, J. R., Swinton, D., and Hattman, S. (1984). 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol *4*, 985-8.

Ramchandani, S., Bigey, P., and Szyf, M. (1998). Genomic structure of the human DNA methyltransferase gene. Biol Chem *379*, 535-540.

Rao, P. M., Antony, A., Rajalakshmi, S., and Sarma, D. S. (1989). Studies on hypomethylation of liver DNA during early stages of chemical carcinogenesis in rat liver. Carcinogenesis *10*, 933-7.

Razin, A., and Riggs, A. D. (1980). DNA methylation and gene function. Science *210*, 604-610.

Razin, A., Szyf, M., Kafri, T., Roll, M., Giloh, H., Scarpa, S., Carotti, D., and Cantoni, G. L. (1986). Replacement of 5-methylcytosine by cytosine: a possible mechanism for transient DNA demethylation during differentiation. Proc Natl Acad Sci U S A *83*, 2827-31.

Razin, A., Webb, C., Szyf, M., Yisraeli, J., Rosenthal, A., Naveh-Many, T., Sciaky-Gallili, N., and Cedar, H. (1984). Variations in DNA methylation during mouse cell differentiation in vivo and in vitro. Proc Natl Acad Sci U S A *81*, 2275-9.

Rideout, W. M., Eversole-Cire, P., Spruck, C. H. r., Hustad, C. M., Coetzee, G. A., Gonzales, F. A., and Jones, P. A. (1994). Progressive increases in the methylation status and heterochromatinization of the myoD CpG island during oncogenic transformation. Mol Cell Biol *14*, 6143-6152.

Rideout, W. M. d., Coetzee, G. A., Olumi, A. F., and Jones, P. A. (1990). 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. Science *249*, 1288-90.

Riggs, A. D. (1990). DNA methylation and late replication probably aid cell memory, and type I DNA reeling could aid chromosome folding and enhancer function. Philos Trans R Soc Lond B Biol Sci *326*, 285-97.

Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet *14*, 9-25.

Roberts, R. J., Myers, P. A., Morrison, A., and Murray, K. (1976). A specific endonuclease from Haemophilus haemolyticus. J Mol Biol *103*, 199-208.

Rouleau, J., Tanigawa, G., and Szyf, M. (1992). The mouse DNA methyltransferase 5'-region. A unique housekeeping gene promoter. J Biol Chem *267*, 7368-7377.

Russell, D. W., and Hirata, R. K. (1989). The detection of extremely rare DNA modifications. Methylation in dam- and hsd- Escherichia coli strains. J Biol Chem *264*, 10787-94.

Russell, L. D., Ettlin, R. A., P., S. H. A., and Clegg, E. D. (1990). Histological and histopathological evaluation of the testis (USA: Cache River Press).

Sakai, T., Toguchida, J., Ohtani, N., Yandell, D. W., Rapaport, J. M., and Dryja, T. P. (1991). Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. Am J Hum Genet *48*, 880-8.

Sanford, J. P., Clark, H. J., Chapman, V. M., and Rossant, J. (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes Dev *1*, 1039-46.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A *74*, 5463-7.

Senapathy, P., Shapiro, M. B., and Harris, N. L. (1990). Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol *183*, 252-78.

Setchell, B. P., Maddocks, S., and Brooks, D. E. (1994). Physiology of reproduction, E. Knobil and J. D. Neill, eds. (New York: Raven Press).

Shapiro, M. B., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res *15*, 7155-74.

Shen, J. C., Rideout, W. M., 3rd, and Jones, P. A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double- stranded DNA. Nucleic Acids Res *22*, 972-6.

Simile, M. M., Pascale, R., De Miglio, M. R., Nufris, A., Daino, L., Seddaiu, M. A., Gaspa, L., and Feo, F. (1994). Correlation between S-adenosyl-L-methionine content and production of c- myc, c-Ha-ras, and c-Ki-ras mRNA transcripts in the early stages of rat liver carcinogenesis. Cancer Lett *79*, 9-16.

Simpson, V. J., Johnson, T. E., and Hammen, R. F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res *14*, 6711-6719.

Steenman, M. J., Rainier, S., Dobry, C. J., Grundy, P., Horon, I. L., and Feinberg, A. P. (1994). Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. Nat Genet *7*, 433-9.

Stener, B., and Peterson, L. (1981). Transacetabular arthrotomy of the hip joint. Acta Orthop Scand *52*, 65-8.

Takagi, H., Tajima, S., and Asano, A. (1995). Overexpression of DNA methyltransferase in myoblast cells accelerates myotube formation. Eur J Biochem *231*, 282-291.

Tate, P., Skarnes, W., and Bird, A. (1996). The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. Nat Genet *12*, 205-8.

Tate, P. H., and Bird, A. P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev *3*, 226-31.

Tollefsbol, T. O., and Andrews, L. G. (1993). Mechanisms for methylation-mediated gene silencing and aging. Med Hypotheses *41*, 83-92.

Tornaletti, S., and Pfeifer, G. P. (1995). Complete and tissue-independent methylation of CpG sites in the p53 gene: implications for mutations in human cancers. Oncogene *10*, 1493-1499.

Toth, M., Muller, U., and Doerfler, W. (1990). Establishment of de novo DNA methylation patterns. Transcription factor binding and deoxycytidine methylation at CpG and non-CpG sequences in an integrated adenovirus promoter. J Mol Biol *214*, 673-683.

Trasler, J. M., Alcivar, A. A., Hake, L. E., Bestor, T., and Hecht, N. B. (1992). DNA methyltransferase is developmentally expressed in replicating and non-replicating male germ cells. Nucleic Acids Res *20*, 2541-2545.

Trasler, J. M., Hake, L. E., Johnson, P. A., Alcivar, A. A., Millette, C. F., and Hecht, N. B. (1990). DNA methylation and demethylation events during meiotic prophase in the mouse testis. Mol Cell Biol *10*, 1828-34.

Trautner, T. A., Balganesh, T. S., and Pawlek, B. (1988). Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. Nucleic Acids Res *16*, 6649-58.

Tucker, K. L., Talbot, D., Lee, M. A., Leonhardt, H., and Jaenisch, R. (1996). Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. Proc Natl Acad Sci USA *93*, 12920-12925.

Turker, M. S. (1990). Methylation of mouse adenine phosphoribosyltransferase gene is altered upon cellular differentiation and loss of phenotypic expression. Somat Cell Mol Genet *16*, 331-40.

Urieli-Shoval, S., Gruenbaum, Y., Sedat, J., and Razin, A. (1982). The absence of detectable methylated bases in Drosophila melanogaster DNA. FEBS Lett *146*, 148-52.

Vairapandi, M., and Duker, N. J. (1993). Enzymic removal of 5-methylcytosine from DNA by a human DNA-glycosylase. Nucleic Acids Res *21*, 5323-7.

Van den Wyngaert, I., Sprengel, J., Kass, S. U., and Luyten, W. H. (1998). Cloning and analysis of a novel human putative DNA methyltransferase. FEBS Lett *426*, 283-289.

Wainfan, E., and Poirier, L. A. (1992). Methyl groups in carcinogenesis: effects on DNA methylation and gene expression. Cancer Res *52*, 2071s-2077s.

Weiss, A., Keshet, I., Razin, A., and Cedar, H. (1996). DNA demethylation in vitro: involvement of RNA. Cell *86*, 709-718.

Wiebauer, K., and Jiricny, J. (1989). In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells. Nature *339*, 234-6.

Wiebauer, K., and Jiricny, J. (1990). Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. Proc Natl Acad Sci U S A *87*, 5842-5.

Wilcox, J. N. (1993). Fundamental principles of in situ hybridization. J Histochem Cytochem *41*, 1725-33.

Wilks, A., Seldran, M., and Jost, J. P. (1984). An estrogen-dependent demethylation at the 5' end of the chicken vitellogenin gene is independent of DNA synthesis. Nucleic Acids Res *12*, 1163-1177.

Willard, H. F., and Hendrich, B. D. (1999). Breaking the silence in Rett syndrome. Nat Genet *23*, 127-8.

Wu, J. C., and Santi, D. V. (1987). Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem *262*, 4778-4786.

Yaffe, D., and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature *270*, 725-7.

Yen, R. W., Vertino, P. M., Nelkin, B. D., Yu, J. J., el-Deiry, W., Cumaraswamy, A., Lennon, G. G., Trask, B. J., Celano, P., and Baylin, S. B. (1992). Isolation and characterization of the cDNA encoding human DNA methyltransferase. Nucleic Acids Res *20*, 2287-91.

Yoder, J. A., and Bestor, T. H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet *7*, 279-284.

Yoder, J. A., Yen, R. W. C., Vertino, P. M., Bestor, T. H., and Baylin, S. B. (1996). New 5' regions of the murine and human genes for DNA (cytosine-5)- methyltransferase. J. Biol. Chem. *271*, 31092-31097.

Zion, M., Ben-Yehuda, D., Avraham, A., Cohen, O., Wetzler, M., Melloul, D., and Ben-Neriah, Y. (1994). Progressive de novo DNA methylation at the bcr-abl locus in the course of chronic myelogenous leukemia. Proc Natl Acad Sci U S A *91*, 10722-6.

A ACKNOWLEDGEMENTS

I wish to express my gratitude to all those who have given me valuable assistance during this project. I am grateful to Drs. M. C. Cardoso and H. Leonhardt for giving me the opportunity to carry out this work. My special thanks for their valuable discussions and criticisms of this work.

There are numerous other people who have assisted me in the course of this project. The group of Dr. M. Bader, especially Drs. T. Walther and M. Reule were very helpful providing me with the mice and their assistance anytime I needed. I would also like to mention the valuable assistance of L. Li during the embryo studies and the technical help provided by Dr. C. Quensel. I also want to mention D. Hänlein for his assistance with computer and printers.

Dr. H. P. Rahn and I. Grunewald deserve special mention for their technical assistance, continuous support and friendship. They also made it possible that this work reached an end. I also have to mention each of my colleagues who in one way or the other have made my work a pleasure. I would like to mention Drs. J. Meding, A. Sportbert and J. B. Margot who always have had a piece of advice at difficult times. It was specially nice to work with J. B. Margot at some stages of this study. Hariharan for the many work discussions and his company during the endless nights in the lab and Dr. W. Derer for his encouragement. I also appreciate the reading and comments of this work by most of my colleagues. Outside science, I would like to mention the many medical doctors in the department who were a constant source of entertaiment with their notorius jokes and sense of humour.

During the course of this work a very special family was always there whenever I needed it. They are the family Boschan, their members Anne, Christoph and Cornelia are just like my own family by now. They did not only have infinitive patience trying to understand my German but they also got involved with my daily life in the lab.

Finally, I would like to thank M. Grupe for his help, advice, support during the many sleepless nights and patience through this undertaking.

Thank you for the support and understanding which you all have shown.

B PERSONAL RECORD

Name: Ana Maria Aguirre-Arteta

Date of birth: 2.01.68, Bilbao (Spain)

Marital status: Single

Nationality: Spanish

EDUCATION

8/1986-6/1991 Bachelor in Technical Biology .

Basque University. Faculty of Biological Sciences.

Lejona. Spain .

2/1993-2/1996 Master in Medical Sciences.

University of Natal. Durban. South Africa.

Title: The relative contribution of the p53 tumor suppressor gene and the human papilloma virus to cancer of the cervix in black women.

Supervisor: Dr RJ Pegoraro. Department of Chemical Pathology. University of Natal. Durban. South Africa.

2/1996-3/2000 PhD student.

Title: Regulation of DNA methylation during development: Alternative isoforms of DNA methyltransferase.

Max Delbrück Center for Molecular Medicine. Franz Volhard Clinic.

Humboldt University . Berlin. Germany.

Supervisors: Drs MC Cardoso and H Leonhardt.

RESEARCH EXPERIENCE

7/ -8/1991 Food and water analysis. Microbiology Department.

Bilbao Public Health. Spain.

5/-9/1992 Analysis of hazardous chemical compounds.

Technical Biology Department. Dublin City University. Glasnevin. Dublin. Ireland.

11/-12/1993 Training in chromatography techniques.

Langet Laboratories. Durban. South Africa.

MEETINGS AND PAPERS PRESENTED

Congresses: * Federation of South African Societies of Pathology 34th Annual Congress, 3-6 July 1994.

University of Cape Town. South Africa.

* Faculty Research Day, 7-Sep. 1994.

Faculty of Medicine. University of Natal. Durban. South Africa.

Title: Loss of heterozygosity of the p53 tumor suppressor gene is not common in cervical cancer in Black South Africans.

Congresses: * Federation of South African Societies of Pathology, 35th Annual Congress, 2-5 July 1995.

Bloemfontein. South Africa.

* Faculty Research Day, 6- Sep. 1995.

Faculty of Medicine. University of Natal. Durban. South Africa.

Title: Confirmation of the low frequency of p53 aberrations in tumors of the cervix.

AWARDS AND HONORS

During my research carried out in Dublin I was awarded the COMETT fellowship.

My Masters studies in South Africa were supported with a Postgraduate Scholarship from the Faculty of Medicine Kennedy Potts Cancer Research Endowment Fund and the Cancer Association of South Africa.

During my PhD studies in Berlin, I was awarded a prize for poster presentation at the 5th Graduate MDC Students' Symposium.

PUBLICATIONS

Chetty R., Bramdev A., Aguirre-Arteta A., Pegoraro R.J., Sataar N. Relation between retinoblastoma and p53 proteins in human papilloma viruses 16/18 positive and negative cancers of the uterine cervix. Journal of Clinical Pathology, 1997: 50 (5): 413-416.

Krämer J., Aguirre-Arteta A.M., Corinna T., Gross C.M., Dietz R., Cardoso M.C., Leonhardt H. A novel isoform of the smooth muscle cell differentiation marker smoothelin. Journal of Molecular Medicine, 1999: 77: 294-298.

*Margot J.B., *Aguirre-Arteta A.M., Di Giacco B.V., Pradham S., Roberts R.J., Cardoso M.C., Leonhardt H. Structure and Function of the Mouse DNA Methyltransferase Gene. Dnmt1 shows a tripartite structure. Journal of Molecular Biology, 2000: 297 (2): 293-300. *Shared first authorship.

Aguirre-Arteta A. M., Cardoso M. C., Leonhardt H. Expression of an alternative Dnmt1 isoform during muscle differentiation. Submitted.

C STATEMENT

Hiermit erkläre ich, die vorliegende Arbeit selbständig angefertig zu haben. Ich habe keine unerlaubten sowie unerwähnten Hilfen benutzt.

Berlin, 31.03.00 Ana M. Aguirre-Arteta